

DIE STATISTISCHE ANALYSE VON
ERGODISCHEN KETTEN MIT
VOLLSTÄNDIGEN BINDUNGEN

Helmut Pruscha

1975



Digitized by the Internet Archive
in 2026

DIE STATISTISCHE ANALYSE VON ERGODISCHEN
KETTEN MIT VOLLSTÄNDIGEN BINDUNGEN

Inaugural-Dissertation
zur Erlangung des Doktorgrades
der Fakultät für Mathematik der
Ludwig-Maximilians-Universität in München

vorgelegt von Helmut Pruscha aus Düsseldorf
im Mai 1974



1. Berichterstatter: Prof. Dr. R. Theodorescu

2. Berichterstatter: Prof. Dr. H. G. Kellerer

Tag der mündlichen Prüfung: 28. Mai 1975

DIE STATISTISCHE ANALYSE VON ERGODISCHEN KETTEN MIT VOLLSTÄNDIGEN BINDUNGEN

	Seite
EINFÜHRUNG	1
BEZEICHNUNGEN	5
TEIL I ZUR THEORIE DER KETTEN MIT VOLLSTÄNDIGEN BINDUNGEN (= KETTEN MVB)	
1. Definitionen und grundlegende Sätze	6
2. OM-Ketten	12
3. Gleichmäßig ergodische Ketten mvB	22
4. Limesverhalten endlicher linearer OM-Ketten	34
TEIL II PARAMETERSCHÄTZUNG IN GLEICHMÄSSIG ERGODISCHEN KETTEN MVB	
5. Dichten für die Übergangswahrscheinlichkeiten $P(s,A)$	39
6. Eine Version des Shannon-McMillan Theorems	49
7. Abhängigkeit der Kette mvB von einem Parameter θ und Konsistenz der Maximum-Likelihood-Schätzung	56
8. Beispiele	68
9. Ein Iterationsverfahren zur Berechnung der Maximum- Likelihood-Schätzung in linearen OM-Ketten	74
ANHANG	82
ZUSAMMENFASSUNG	85
LITERATURVERZEICHNIS	89

EINFÜHRUNG

Zur Beschreibung des Ablaufes biologischer Vorgänge reicht das Modell einer Markovkette im allgemeinen nicht aus, obwohl man es in der Anwendung fast ausschließlich findet, wohl auf Grund seiner einfachen statistischen Handhabung (die relativen Übergangshäufigkeiten n_{ij}/n_i sind einfach zu gewinnende konsistente Schätzungen der zugrundeliegenden Übergangswahrscheinlichkeiten Λ_{ij} , wenn Λ eine geeignete stochastische Matrix ist).

Ist $(\xi_n)_{n \in \mathbb{N}}$ eine (homogene) Markovkette mit Werten in einem - für den Augenblick als endlich vorausgesetzt - Ereignisraum X und mit der Matrix Λ der Übergangswahrscheinlichkeiten, so ist die Wahrscheinlichkeitsverteilung für ξ_{n+1} unter der Bedingung, daß vorher die Ereignisse $\xi_1=x_1, \dots, \xi_n=x_n$ auftraten

$$P(\xi_{n+1}=x | x_1, \dots, x_n) = \Lambda_{x_n x},$$

wobei x den Ereignisraum X durchläuft. Das Auftreten eines jeden Ereignisses löscht also die Vorgeschichte aus und setzt vollständig die Wahrscheinlichkeiten für das nächstfolgende Ereignis fest. Diese Einschränkung eliminiert aber gerade wesentliche Eigenschaften biologischer Vorgänge.

Eine Erweiterung der Markovkette mit wenigen zusätzlichen Parametern ist z.B. die sogenannte lineare OM-Kette (vgl. Def. (2.6))

$$(*) \quad P(\xi_{n+1}=x | x_1, \dots, x_n) = \alpha_{x_n} P(\xi_n=x | x_1, \dots, x_{n-1}) + (1-\alpha_{x_n}) \Lambda_{x_n x},$$

$|\alpha_{x_n}| \leq 1$, in der die Wahrscheinlichkeitsverteilung von ξ_{n+1}

im allgemeinen von der gesamten Vorgeschichte $x_1, \dots, x_n$ abhängt. Die lineare OM-Kette bringt für die einzelnen Ereignisse $x \in X$ durch die Koeffizienten α_x eine zusätzliche Charakterisierung: Die Wahrscheinlichkeiten für das nächstfolgende Ereignis werden mit dem Gewicht $(1-\alpha_{x_n})$ durch das n -te Ereignis x_n und mit dem Gewicht α_{x_n} durch die Vorgeschichte $x_1, \dots, x_{n-1}$ bestimmt.

Jedoch ist ein Ansatz wie in (*) nur für die Anwendung brauchbar, wenn man weiß, wie man aus einer Stichprobe $x_1, x_2, \dots, x_n$ eine derartige Schätzung $\hat{\alpha}_x$ für die α_x gewinnt, daß die $\hat{\alpha}_x$ bei $n \rightarrow \infty$ nach Wahrscheinlichkeit gegen die zugrundeliegenden α_x konvergieren. Die Frage nach einer solchen konsistenten Schätzung für die α_x motivierte diese Arbeit. Zweckmäßigerweise betrachtet man die Kette (*) als Spezialfall einer sogenannten Kette mit vollständigen Bindungen (Kette mvB) mit beliebigem Ereignisraum X . Denn erstens ist die zugehörige Theorie - abgesehen von einer statistischen Theorie - sehr weit entwickelt (vgl./13/), und zum zweiten stellt der Begriff der Kette mvB auch allgemeinere Modelle als (*) zur Beschreibung biologischer Vorgänge bereit.

Im Teil I der Arbeit wird nun die vorhandene Theorie der Ketten mvB und speziell der linearen OM- Ketten gemäß unseren Bedürfnissen ergänzt.

Im Teil II der Arbeit führen wir für die Übergangswahrscheinlichkeiten einer Kette mvB Dichten ein, die noch von einem unbekannten Parameter $\vartheta \in \mathbb{R}^m$ abhängen. Es wird dann gezeigt, daß unter verschiedenen Voraussetzungen die Maximum-Likelihood-Schätzung für den wahren Parameterwert ϑ_0 eine konsistente Schätzung ist (Satz (7.15)).

Das Problem, die Konsistenz der Maximum-Likelihood-Schätzung in einem stochastischen Prozeß mit diskretem Zeitparameter zu beweisen, hat zwei Traditionen.

Die erste Tradition geht von der Forderung $\partial \log L(\vartheta) / \partial \vartheta_1 = 0$ aus ($L(\vartheta)$ sei die Likelihoodfunktion einer Stichprobe in Abhängigkeit von ϑ) und entwickelt die linke Seite der Gleichung gemäß

Taylor im Punkte ϑ_0 . Arbeiten in dieser Richtung liegen z.B. vor von Wald /26/ für eine Kette unabhängiger, identisch verteilter Variablen, Billingsley /5/ für ergodische Markovprozesse, Rao /22/ und Weiss /28/ für beliebige stochastische Prozesse, allerdings mit Vorschriften über das Limesverhalten von Ausdrücken, in denen diverse Ableitungen von $L(\vartheta)$ und $\log L(\vartheta)$ eingehen. Die zweite Tradition geht direkt von der Forderung $\log L(\vartheta) = \text{Maximal}$ aus, ohne auf die Differenzierbarkeit von $\log L(\vartheta)$ nach ϑ einzugehen, und benutzt an zentraler Stelle das individuelle Ergodentheorem, um die Konvergenz von $\frac{1}{n} \log L(\vartheta)$ P-fast überall zu zeigen. Man erhält dann sogar "starke" Konsistenz, d.h. die Konvergenz der Schätzung P-fast überall. In dieser Richtung arbeiteten Wald /27/ für Ketten unabhängiger, identisch verteilter Zufallsvariablen, Baum und Petrie /2/ und Petrie /20/ für probabilistische Funktionen ergodischer Markovketten und Roussas /23/, Gänsler /11/ für ergodische Markovprozesse. Die vorliegende Arbeit schließt sich dieser zweiten Tradition an, weil die umfangreichen Voraussetzungen über die Ableitungen von $\log L(\vartheta)$ bereits für den Fall (*) der endlichen linearen OM-Ketten kaum nachprüfbar sind.

Gegenüber den oben angeführten Arbeiten mußte im Teil II dieser Arbeit die Beweisführung so abgewandelt und erweitert werden, daß die allgemeinere Art der stochastischen Abhängigkeit und vor allem die Nichtstationarität einer Kette mvB bewältigt werden konnten.

Schließlich bleibt noch die Frage, wie man die Stelle praktisch berechnen soll, an der die Likelihoodfunktion $L(\vartheta)$ ihr Maximum annimmt. Dazu wird ein durch /3/ inspiriertes Iterationsverfahren vorgestellt, das den Maximum-Likelihood-Schätzwert für die Parameter der endlichen linearen OM-Kette berechnet.

Angeregt zu dieser Arbeit wurde ich durch einen Vortrag von Herrn Professor Theodorescu im Mathematischen Institut der Universität München im Dezember 1971, in welchem dieser auf die offenen Probleme der statistischen Theorie für Ketten mit vollständigen Bindungen hinwies. Die vorliegende Arbeit will einen Beitrag zu dieser Theorie geben.

BEZEICHNUNGEN

Ist X eine Menge, so bezeichnet X^n das n -fache kartesische Produkt dieser Menge und $x^n = (x_1, \dots, x_n)$ ein Element von X^n .

Ist $\mathcal{K}$ eine σ -Algebra, so bezeichnet die σ -Algebra $\mathcal{K}^n$ das n -fache Produkt von $\mathcal{K}$ und A^n ein Element von $\mathcal{K}^n$, wobei nicht notwendig $A^n = (A_1, \dots, A_n)$, $A_i \in \mathcal{K}$, angenommen wird.

Mit der Menge $\mathbb{N}$ der natürlichen Zahlen und der Menge $\mathbb{Z}$ der ganzen Zahlen sind

$$X^{\mathbb{N}}, X^{\mathbb{Z}}, x^{\mathbb{N}}, x^{\mathbb{Z}}, \mathcal{K}^{\mathbb{N}}, \mathcal{K}^{\mathbb{Z}}, A^{\mathbb{N}}, A^{\mathbb{Z}}$$

analog definiert.

Ist $(X, \mathcal{K})$ ein Maßraum und sind $\xi_1, \dots, \xi_n$ meßbare Funktionen auf $(X, \mathcal{K})$ [bzw. $\mathcal{N}$ ein Mengensystem aus $\mathcal{K}$], so bezeichne $\sigma(\xi_1, \dots, \xi_n)$ [bzw. $\sigma(\mathcal{N})$] die von den $\xi_1, \dots, \xi_n$ [bzw. von $\mathcal{N}$] erzeugte Unter- σ -Algebra von $\mathcal{K}$.

Für ein $n \in \mathbb{Z}$ bezeichne $(-\infty, n]$ die Menge $\{\dots, n-1, n\} \subset \mathbb{Z}$ und $[n, \infty)$ die Menge $\{n, n+1, \dots\} \subset \mathbb{Z}$.

Für $n, m \in \mathbb{Z}$, $n < m$, bezeichne $[n, m]$ die Menge $\{n, n+1, \dots, m-1, m\} \subset \mathbb{Z}$.

Dementsprechend hat man

$X^{(-\infty, n]}$, $X^{[n, \infty)}$, $X^{[n, m]}$ usw. zu verstehen.

TEIL I Zur Theorie der Ketten mit vollständigen Bindungen

1. Definitionen und grundlegende Beziehungen

Zur Definition einer Kette mit vollständigen Bindungen (in Zukunft nur noch "Kette mvB" genannt) übernehmen wir das Konzept von Iosifescu und Theodorescu /13/, Seite 63.

(1.1) Definition Ein System $((X, \mathcal{X}), (S, \mathcal{S}), P, u)$ heißt Kette mvB, falls folgende Bedingungen erfüllt sind:

- (a) $(X, \mathcal{X}), (S, \mathcal{S})$ sind Maßräume
- (b) $P: S \times \mathcal{X} \longrightarrow [0, 1]$ ist ein Markovkern von $(S, \mathcal{S})$ nach normiertes
 $(X, \mathcal{X})$, d.h.: für jedes $s \in S$ ist $P(s, \cdot)$ ein $\mathcal{X}$ -Maß auf $\mathcal{X}$
 und für jedes $A \in \mathcal{X}$ ist $P(\cdot, A)$ $\mathcal{S}$ -meßbar
- (c) $u: S \times X \rightarrow S$ ist eine $\mathcal{S} \times \mathcal{X}$ - $\mathcal{S}$ -meßbare Abbildung.

Für jedes $n \in \mathbb{N}$ und $x^n = (x_1, x_2, \dots, x_n) \in X^n$ definiert man dann die $(\mathcal{S} \times \mathcal{X}^n - \mathcal{S}$ -meßbare) iterierte Abbildung

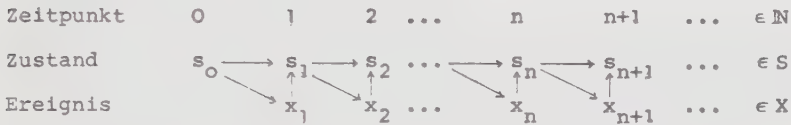
$$u(s, x^n) := u(u(\dots u(u(s, x_1), x_2) \dots x_{n-1}), x_n)$$

(1.2) Bemerkung Im Sinne von /13/ handelt es sich hier bereits um eine homogene Kette mvB, weil die Abbildungen P und u nicht von einem $n \in \mathbb{N}$ oder $\mathbb{Z}$ abhängen. Wir wollen das in (1.1) definierte System auch kurz mit (X, S, P, u) bezeichnen. Ist X eine endliche Menge, so spricht man von einer endlichen Kette mvB.

(1.3) Interpretation und Ablauf einer Kette mvB

Wir nennen S die Menge der Zustände ("state space" in /13/, S. 203) und X die Menge der Ereignisse ("event space"). Stellt man sich für einen Moment X als endliche Menge vor, so läßt

sich zu einer Kette mvB folgender zeitlicher Ablauf
 assoziieren: Zum Zeitpunkt 0 existiert ein Anfangszustand
 $s_0 \in S$. Mit Wahrscheinlichkeit $P(s_0, x_1)$ tritt zum Zeit-
 punkt 1 das Ereignis x_1 ein, wodurch sich der neue Zustand
 $s_1 = u(s_0, x_1)$ einstellt. Mit Wahrscheinlichkeit $P(s_1, x_2)$
 tritt zum Zeitpunkt 2 das Ereignis x_2 ein, usw. Graphisch
 dargestellt:



Das folgende Existenztheorem von Iosifescu und Theodorescu
 zeigt, daß eine Kette mvB auf einem Wahrscheinlichkeitsraum
 durch einen stochastischen Prozeß realisiert werden kann.

(1.4) Satz Sei (X, S, P, u) eine Kette mvB. Zu jedem $s \in S$
 existiert ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{H}, P_s)$ zusammen
 mit einer Folge (ξ_n) von Zufallsvariablen auf Ω mit Werten
 in X , so daß

$$P_s(\xi_1 \in A) = P(s, A) \text{ und}$$

$$P_s(\xi_{n+1} \in A \mid \sigma(\xi^n)) = P(u(s, \xi^n), A) \quad P_s\text{-fast überall}$$

wobei $n \in \mathbb{N}$, $A \in \mathcal{X}$, $\xi^n = (\xi_1, \dots, \xi_n)$

Beweis /13/, Seite 64. Dort wird gesetzt

$$\Omega := X^{\mathbb{N}}, \quad \mathcal{H} := \mathcal{X}^{\mathbb{N}},$$

$$\xi_n(x^{\mathbb{N}}) := x_n \text{ für } x^{\mathbb{N}} \in X^{\mathbb{N}}, n \in \mathbb{N}$$

$$P_s(\xi_1 \in A) := P(s, A) \text{ für } A \in \mathcal{X}$$

$$\text{und für } n > 1, A_1, A_2, \dots, A_n \in \mathcal{X}$$

$$P_s(\xi_1 \in A_1, \dots, \xi_n \in A_n) :=$$

$$\int_{A_1} P(s, dx_1) \int_{A_2} P(u(s, x_1), dx_2) \dots \int_{A_{n-1}} P(u(s, x^{n-2}), dx_{n-1}) \cdot P(u(s, x^{n-1}), A_n)$$

Die Existenz einer Wahrscheinlichkeit P_s auf $\mathcal{X}^{\mathbb{N}}$ wird nun durch ein Theorem von Ionescu Tulcea (vgl. /16/, Prop. V 1.1) und die im Satz behauptete Beziehung durch die Definition der bedingten Wahrscheinlichkeit garantiert.

Eine Umkehrung von (1.4) gilt unter der zusätzlichen topologischen Voraussetzung, daß X ein polnischer Raum ist (polnischer Raum = separabler, vollständig metrisierbarer Raum). In /13/, S. 86, ist die Umkehrung ohne Beweis angedeutet.

(1.5) Satz Sei $(\Omega, \mathcal{K}, \tilde{P})$ ein Wahrscheinlichkeitsraum und (ξ_n) eine Folge von Zufallsvariablen auf Ω mit Werten im polnischen Raum X . $\mathcal{K}$ seien die Borelschen Mengen von X . Dann existiert eine Kette mvB (X, S, P, u) und ein $s_0 \in S$, so daß

$$P(s_0, A) = \tilde{P}(\xi_1 \in A)$$

$$P(u(s_0, \xi^n), A) = \tilde{P}(\xi_{n+1} \in A \mid \sigma(\xi^n)) \quad [\tilde{P}\text{-fast überall}],$$

wobei $n \in \mathbb{N}$, $A \in \mathcal{K}$ ist.

Beweis Setze $S := \bigcup_{n \in \mathbb{N} \cup \{0\}} X^n$, $\mathcal{K} := \sigma(\bigcup_{n \in \mathbb{N} \cup \{0\}} \mathcal{K}^n) =$

$\mathcal{K}^0 \times \mathcal{K}^{\mathbb{N}}$, wobei $X^0 = \{x^0\}$ eine beliebige, aber zu den

X^n , $n \in \mathbb{N}$, disjunkte, einpunktige Menge sei,

$$s_0 := x^0, u(s_0, x) := x \text{ und } u(s, x) = (x_1, \dots, x_n, x), \text{ falls}$$

$$s = (x_1, \dots, x_n).$$

$u(s, x)$ ist klarerweise $\mathcal{K} \times \mathcal{K}$ - $\tilde{\mathcal{K}}$ -meßbar.

$$\text{Setze weiter } P(s_0, A) := \tilde{P}(\xi_1 \in A).$$

Nach einem Satz über bedingte Verteilungen (/1/ , Satz 56.5) existiert eine solche ($\tilde{P}$ -fast eindeutige) Festlegung der bedingten Wahrscheinlichkeit $\tilde{P}(\xi_{n+1} \in A \mid \mathfrak{F}(\xi^n))$, die Markovkern von $(X^n, \mathfrak{G}(\xi^n))$ nach $(X, \mathfrak{K})$ ist. Deshalb ist $P(s, A) := \tilde{P}(\xi_{n+1} \in A \mid (\xi_1 = x_1, \dots, \xi_n = x_n))$ für $s = (x_1, \dots, x_n)$ ein Markovkern von $(S, \mathfrak{U})$ nach $(X, \mathfrak{K})$. Weil nach Definition $u(s_0, x^n) = x^n$ gilt, folgt wie behauptet: $P(u(s_0, \xi^n), A) = \tilde{P}(\xi_{n+1} \in A \mid \mathfrak{G}(\xi^n))$.

(1.6) Bemerkung Auf einem polnischen Raum hat man demnach die Begriffe stochastischer Prozeß mit diskretem Zeitparameter N und Kette mvB als äquivalent anzusehen. Trotzdem ist eine eigenständige Theorie der Ketten mvB vorteilhaft, weil sie zu Modellbildungen für die Abhängigkeit zukünftiger Ereignisse von der Vergangenheit anregt, auf die man vom Begriff des stochastischen Prozesses her nicht kommen würde. Letzterer hat nämlich für das Studium allgemeiner Verhaltensabläufe im wesentlichen nur das Modell der Markovkette (beliebiger Ordnung) bereitgestellt. Das Markovmodell wird sich aber als Sonderfall einer speziellen Kette mvB ergeben, nämlich der linearen OM-Kette. (siehe (2.12))

Mit Hilfe von Satz (1.4) kann man die Definition der Wahrscheinlichkeit $P(s, \cdot) \mid \mathfrak{K}$ ausdehnen auf $P(s, \cdot) \mid \mathfrak{K}^{\mathbb{N}}$:

(1.7) Definition Für $A^{\mathbb{N}} \in \mathfrak{K}^{\mathbb{N}}$ sei

$P(s, A^{\mathbb{N}}) := P_s((\xi^{\mathbb{N}}) \in A^{\mathbb{N}})$, und für $k \in \mathbb{N}$ sei

$P^k(s, A^{\mathbb{N}}) := P(s, T^{-(k-1)} A^{\mathbb{N}})$ die sogenannte k -Schritt-

Wahrscheinlichkeit von $P(s, \cdot)$ [$T: X^{\mathbb{N}} \rightarrow X^{\mathbb{N}}$ ist def.

durch $(T x^{\mathbb{N}})_n = x_{n+1}$, vgl. (3.11)].

(1.8) Lemma Für $A^n \in \mathfrak{K}^n$, $n \in \mathbb{N}$, gilt

$$P(s, A^n) = \int_{A^n} P(s, dx_1) \cdot P(u(s, x_1), dx_2) \dots P(u(s, x^{n-1}), dx_n)$$

Beweis Nach Definition ist $P(s, A^n) = P_s((\xi^n) \in A^n)$.

Die Behauptung folgt aus der Festlegung der Wahrscheinlichkeit P_s im Beweis zu Satz (1.4).

Im Schaubild zu (1.3) ist die Folge $s_0, s_1, s_2, \dots$ von Zuständen eingetragen. Sie bilden, wie man sich überlegen kann, eine einfache Markovkette (einfach = der Ordnung 1).
Genauer gilt:

(1.9) Satz Sei (X, S, P, u) eine Kette mvB, $s_0 \in S$, und

$(\Omega, \mathcal{K}, P_{s_0}, (\xi_n))$ der gemäß Satz (1.4) zugehörige stochastische Prozeß. Durch $\zeta_n := u(s_0, \xi_n)$, $n \in \mathbb{N}$, wird eine

Folge von Zufallsvariablen auf $(\Omega, \mathcal{K}, P_{s_0})$ mit Werten in $(S, \mathcal{S})$ definiert. Es gilt: $(\zeta_n)_{n \in \mathbb{N}}$ ist eine einfache Markovkette mit einer Startverteilung, die auf s_0 konzentriert ist, und mit den Übergangswahrscheinlichkeiten

$Q(s, B) := P(s, B_s)$, wobei $B_s = \{x \in X : u(s, x) \in B\} \in \mathfrak{K}$, für $s \in S$ und $B \in \mathcal{S}$.

Beweis /13/, S. 66.

(1.10) Bemerkung

Die entsprechenden k-Schritt-Übergangswahrscheinlichkeiten für die Kette (ξ_n) heißen:

$$Q^k(s, B) = P(s, B_S^k), \text{ wobei } B_S^k = \{x^k \in X^k : u(s, x^k) \in B\} \in \mathcal{X}^k.$$

(1.11) Lemma Für $\mathcal{U}$ -meßbares $\phi(s)$ ist

$$\int_S Q(s, ds') \phi(s') = \int_X P(s, dx) \phi(u(s, x))$$

falls eines der beiden Integrale existiert.

Beweis Setzt man $\phi(s) = \chi_B(s)$, $B \in \mathcal{U}$,

wobei χ_B die Indikatorfunktion der Menge B ist, so gilt die Behauptung aufgrund der Definition von Q. Wie üblich setzt man den Beweis dann fort.

2. OM-Ketten

Der Typ von Ketten mvB, der nun eingeführt wird, wurde zuerst von Onicescu und Mihoc /19/ untersucht, und zwar für endliches X . Er wird deshalb OM-Kette genannt. Iosifescu und Theodorescu /13/, S. 184, erweiterten ihn auf beliebiges X im Sinne der Definition (2.1). Bei OM-Ketten besteht die Zustandsmenge S aus Wahrscheinlichkeitsmaßen auf $\mathcal{X}$.

(2.1) Definition Eine Kette mvB $((X, \mathcal{X}), (S, \mathcal{O}), P, u)$ wird OM-Kette genannt, falls

- (a) $(X, \mathcal{X})$ Meßraum, S eine Menge von Wahrscheinlichkeitsmaßen auf $\mathcal{X}$ und $\mathcal{O}$ eine σ -Algebra über S .
- (b) $P(\mu, A) := \mu(A)$, wobei $\mu \in S$ und $A \in \mathcal{X}$.

Man beachte, daß die Abbildungen P und u natürlich die Meßbarkeitseigenschaften (1.1) b und c erfüllen müssen; vergleiche dazu (2.5), (2.9).

Wir zeigen nun, daß OM-Ketten nicht sehr viel spezieller sind als Ketten mvB:

(2.2) Satz Eine Kette mvB (X, S, P, u) erfülle die Voraussetzungen

- (A₁) $P(s, A) = P(s', A)$ für alle $A \in \mathcal{X} \implies P(u(s, x), A) = P(u(s', x), A)$ für alle $A \in \mathcal{X}$ und $x \in X$.
- (A₂) $P(s, A) = P(s', A)$ für alle $A \in \mathcal{X} \implies$
Es existiert kein $B \in \mathcal{O}$ mit $s \in B, s' \notin B$

Sei $s_0 \in S$. Dann existiert eine OM-Kette $(X, \tilde{S}, \tilde{P}, \tilde{u})$ und ein $\tilde{s}_0 \in \tilde{S}$, so daß die zugehörigen Wahrscheinlichkeitsräume $(\Omega, \mathcal{K}, P_{s_0})$ und $(\Omega, \mathcal{K}, \tilde{P}_{\tilde{s}_0})$ identisch sind.

Beweis

1. Man definiert eine Abbildung Π von S in die Menge der Wahrscheinlichkeitsmaße auf $\mathcal{X}$ vermöge:

$$\Pi : s \longmapsto P(s, \cdot)$$

und setzt $\tilde{S} := \Pi(S)$, $\tilde{\mathcal{O}} := \Pi(\mathcal{O})$.

Weil Π - aufgefaßt als Mengenabbildung - gemäß (A_2) injektiv auf $\mathcal{O}$ ist, bildet $\Pi(\mathcal{O})$ eine σ -Algebra über $\Pi(S)$.

2. Für $\mu \in \tilde{S}$ und $A \in \mathcal{A}$ setzt man

$$\tilde{P}(\mu, A) := \mu(A).$$

Für jedes μ ist $\tilde{P}(\mu, \cdot)$ ein Maß auf $\mathcal{A}$, und für jedes $A \in \mathcal{A}$ ist $\tilde{P}(\cdot, A) : \tilde{\mathcal{O}} \rightarrow \mathbb{R}_1$ -meßbar. Denn für ein Borelsches C ist $B := \{s : P(s, A) \in C\}$ voraussetzungsgemäß aus $\mathcal{O}$ und damit $\{\mu \in \tilde{S} : \tilde{P}(\mu, A) \in C\} = \{\mu \in \tilde{S} : \mu(A) \in C\} = \Pi\{s \in S : P(s, A) \in C\} = \Pi(B) \in \tilde{\mathcal{O}}$.

3. Für ein $\mu \in \tilde{S}$ mit $\mu = \Pi(s)$ und für $x \in X$ setzt man

$$\tilde{u}(\mu, x) := \Pi(u(s, x)).$$

Diese Definition ist eindeutig, denn für ein $s' \in S$ mit

$$\mu = \Pi(s') \text{ gilt } P(s, \cdot) = P(s', \cdot), \text{ und deshalb auch}$$

$$P(u(s, x), \cdot) = P(u(s', x), \cdot) \text{ wegen } (A_1), \text{ woraus}$$

$$\Pi(u(s, x)) = \Pi(u(s', x)) \text{ folgt.}$$

Für jedes $\tilde{B} \in \tilde{\mathcal{O}}$ ist

$$\{(\mu, x) : \tilde{u}(\mu, x) \in \tilde{B}\} \in \tilde{\mathcal{O}} \times \mathcal{X}.$$

Denn für ein $\tilde{B} = \Pi(B)$, $B \in \mathcal{O}$, und mit der Bezeichnung

$$(\Pi, \text{Id}) \text{ für die Produktabbildung } (\Pi, \text{Id})(s, x) := (\Pi(s), x)$$

gilt:

$$\{(\mu, x) : \tilde{u}(\mu, x) \in \tilde{B}\}$$

$$= (\Pi, \text{Id}) \{ (s, x) : \Pi(u(s, x)) \in \tilde{B} \}$$

$$= (\Pi, \text{Id}) \{ (s, x) : u(s, x) \in B \}$$

Die in den Mengenklammern stehende Menge (sie soll mit C_0 abgekürzt werden) liegt in $\mathcal{O} \times \mathcal{X}$ nach (1.1) c.

Um nun zu zeigen, daß (π, Id) (C_0) in $\tilde{\mathcal{O}} \times \mathcal{X}$ liegt, betrachten wir die Menge

$$\mathcal{L} := \{ C \in \mathcal{O} \times \mathcal{X} : (\pi, Id)(C) \in \tilde{\mathcal{O}} \times \mathcal{X} \}$$

$\mathcal{L}$ ist eine σ -Algebra, wenn (π, Id) injektiv auf $\mathcal{O} \times \mathcal{X}$ ist. Dies ist aber wegen Voraussetzung (A_2) der Fall.

$\mathcal{L}$ enthält insbesondere alle Mengen (B, A) , $B \in \mathcal{O}$, $A \in \mathcal{X}$. Es folgt $\mathcal{L} = \mathcal{O} \times \mathcal{X}$ und damit $(\pi, Id)(C_0) \in \tilde{\mathcal{O}} \times \mathcal{X}$.

4. Nun zeigen wir, daß die Wahrscheinlichkeiten P_{s_0} und $\tilde{P}_{\tilde{s}_0}$ auf dem Raum $(X^{\mathbb{N}}, \mathcal{X}^{\mathbb{N}})$ identisch sind. Setze $\tilde{s}_0 := \pi(s_0)$. Dann gilt für $A \in \mathcal{X}$

$$(*) \quad P(s_0, A) = \tilde{P}(\tilde{s}_0, A)$$

$$(**) \quad P(u(s_0, x^n), A) = \tilde{P}(\tilde{u}(\tilde{s}_0, x^n), A) \text{ für } x^n \in X^n.$$

Die erste Gleichung folgt unmittelbar aus den Definitionen.

Die zweite Gleichung gilt, weil aus der Festsetzung

$$\tilde{u}(\tilde{s}, x) = \pi(u(s, x)) \text{ für } \tilde{s} = \pi(s)$$

$$\tilde{u}(\tilde{s}, x^n) = \pi(u(s, x^n)) \text{ folgt.}$$

Beweis durch Induktion:

$$\tilde{u}(\tilde{s}, x^{n+1}) \stackrel{(1.1)c}{=} \tilde{u}(\tilde{u}(\tilde{s}, x^n), x_{n+1}) \stackrel{(\text{Ind.vor.})}{=} \tilde{u}(\pi(u(s, x^n)), x_{n+1})$$

$$\stackrel{(\text{Def.})}{=} \pi(u(u(s, x^n), x_{n+1})) \stackrel{(1.1)c}{=} \pi(u(s, x^{n+1}))$$

Aus (*) und (**) folgt aber wegen Satz (1.4) bereits:

$$P_{s_0} = \tilde{P}_{\tilde{s}_0}.$$

(2.3) Bemerkungen zu den Voraussetzungen (A_1) , (A_2) .

Zu (A_1) : Besonders anschaulich wird (A_1) , wenn man auf die im Satz (1.5) gegebene Darstellung eines stochastischen

Prozesses als Kette mvB zurückgreift:

Gilt für $(x_1, \dots, x_n) \in X^n$ und für $(x'_1, \dots, x'_m) \in X^m$:

$$P((x_1, \dots, x_n), \cdot) = P((x'_1, \dots, x'_m), \cdot)$$

als Wahrscheinlichkeitsmaße auf $\mathcal{X}$, so soll für jedes $x \in X$ folgen:

$$P((x_1, \dots, x_n, x), \cdot) = P((x'_1, \dots, x'_m, x), \cdot)$$

Es fällt schwer, sich einen in der Natur vorkommenden stochastischen Prozeß vorzustellen, bei dem diese Voraussetzung verletzt wird.

Zu (A_2) : (A_2) ist zum Beispiel erfüllt, wenn $P(s, \cdot) \neq P(s', \cdot)$ für $s \neq s'$ als Wahrscheinlichkeitsmaße auf $\mathcal{X}$, oder allgemeiner, wenn $\mathcal{O} = \bigcap_{n=1}^{\infty} (\mathcal{O}_n)$ für irgendeine σ -Algebra $\mathcal{M}$ über der Menge der Wahrscheinlichkeitsmaße auf $\mathcal{X}$.

In Anschluß an Definition (2.1) stellt sich die Frage, für welches $\mathcal{O}$ die Meßbarkeitsbedingung (1.1) b erfüllt ist.

Diese lautet für OM-Ketten:

Die Abbildung $\mu \rightarrow \mu(A)$ ist für jedes $A \in \mathcal{A} \quad \mathcal{O}$ -meßbar.

Dazu führen wir auf der Menge der Wahrscheinlichkeitsmaße die sogenannte starke Topologie ein.

(2.4) Erläuterungen M^e bezeichne die Menge der allgemeinen, endlichen Maße auf $\mathcal{X}$, M die Untermenge der Wahrscheinlichkeitsmaße auf $\mathcal{X}$. Durch die Norm (vgl./12/, (20.43), (19.62), (19.7))

$$\|\mu\| = \frac{1}{2} \left\{ \sup_{A \in \mathcal{A}} \mu(A) - \inf_{A \in \mathcal{A}} \mu(A) \right\}$$

wird auf M^e die starke Topologie induziert. Fassen wir M als topologischen Unterraum von M^e auf, so wird die auf M induzierte starke Topologie auch durch die Metrik

$$d(\mu, \nu) = \sup_{A \in \mathcal{A}} |\mu(A) - \nu(A)|, \quad \mu, \nu \in M$$

erzeugt. Denn ist $\mu(X) = 0$ für ein $\mu \in M^e$, so wird

$$\|\mu\| = \sup_{A \in \mathcal{X}} |\mu(A)|, \text{ wie man nachrechnet.}$$

In der Metrik d ist M sogar vollständig, denn für eine Cauchyfolge $\mu_n \in M$ existiert $\lim_n \mu_n(A)$ für jedes $A \in \mathcal{X}$ (sogar gleichmäßig in $A \in \mathcal{X}$). Die Gleichmäßigkeit der Konvergenz liefert dann, daß $\mu(A) := \lim_n \mu_n(A)$, $A \in \mathcal{X}$, ein Wahrscheinlichkeitsmaß auf $\mathcal{X}$ ist.

(2.5) Lemma Sei $((X, \mathcal{X}), (S, \mathcal{O}), P, u)$ eine OM-Kette, für die $\mathcal{O}$ die Menge der zur starken Topologie gehörenden Borelschen Teilmengen von S ist. Dann ist für jedes $A \in \mathcal{X}$ die Abbildung $\mu \longrightarrow P(\mu, A) = \mu(A)$ $\mathcal{O}$ -meßbar.

Beweis Sei $\mu, \nu \in S$. Aus

$$|\mu(A) - \nu(A)| \leq \sup_{A \in \mathcal{X}} |\mu(A) - \nu(A)| = d(\mu, \nu)$$

folgt, daß die Abbildung $\mu \longrightarrow \mu(A)$ stetig ist für jedes $A \in \mathcal{X}$, also erst recht meßbar.

In der Definition (2.1) wurde bei einer Kette $mvB(X, S, P, u)$ die Zustandsmenge S und die Übergangswahrscheinlichkeit P spezialisiert zu einer OM-Kette. Durch Spezialisierung der Abbildung $u : S \times X \longrightarrow S$ kommen wir zur linearen OM-Kette (vgl. etwa /4/, S. 13 und 14).

(2.6) Def. Eine OM-Kette (X, S, P, u) heißt lineare OM-Kette, falls für jedes $\mu \in S$, $x \in X$ gilt: $u(\mu, x) = \alpha_x \mu + (1 - \alpha_x) \Lambda_x$ mit $\Lambda_x \in S$ (also W.maß auf $\mathcal{X}$) und $0 \leq \alpha_x \leq 1$.

Bemerkung Bei endlichem $\mathcal{X}$ könnten auch negative α_x zugelassen werden (vgl. (2.7)). Wir fordern aber hier $\alpha_x \geq 0$, weil wir dies in den meisten Fällen benötigen (Ausnahmen: (2.9), (3.2), (4.2)).

In Satz (2.2) zeigten wir, daß unter den Voraussetzungen (A_1) , (A_2) eine Kette mvB zu einer OM-Kette wird. Wann stellt nun eine OM-Kette eine lineare OM-Kette dar? Bush, Mosteller und Thompson /7/ bewiesen für OM-Ketten mit endlichem X , daß dies unter der sogenannten "combining of classes condition" der Fall ist. Allgemeiner lautet ihr Ergebnis (M wie in (2.4), $\mu|_{\mathcal{A}}$ bedeutet die Restriktion des Maßes μ auf $\mathcal{A} \subset \mathcal{K}$).

(2.7) Satz Sei (X, M, P, u) eine OM-Kette. Für jede Unter- σ -Algebra $\mathcal{A} \subset \mathcal{K}$ gelte

(A_3) $\mu, \hat{\mu} \in M$ mit $\mu|_{\mathcal{A}} = \hat{\mu}|_{\mathcal{A}} \implies u(\mu, x)|_{\mathcal{A}} = u(\hat{\mu}, x)|_{\mathcal{A}}$ für alle $x \in X$.

Dann gilt $u(\mu, x) = \alpha_x \mu + (1 - \alpha_x) \Lambda_x$ mit $\Lambda_x \in M$ und $-\frac{1}{N-1} \leq \alpha_x \leq 1$ falls $|X| = N$, $\mathcal{K} = \mathcal{P}(X)$; $0 \leq \alpha_x \leq 1$ falls $|\mathcal{K}| = \infty$.

Beweis folgt /7/, Theorem 1, p.124. Anstelle der dort auftretenden Zerlegungen des endlichen Raumes X bildet man geeignete Unter- σ -Algebren von $\mathcal{K}$, wie im ANHANG ausgeführt ist.

In Anschluß an Definition (2.6) stellt sich wieder die Frage, wann die Meßbarkeitsvoraussetzung (1.1) c erfüllt ist.

(2.8) Erläuterungen Für ein σ -endliches Maß $1|\mathcal{K}$ bezeichne M_1 die Menge aller Wahrscheinlichkeitsmaße über $\mathcal{K}$, die eine 1-Dichte besitzen. Es ist $M_1 \subset M$. Sind f_μ bzw. f_ν die 1-Dichten für μ bzw. $\nu \in M_1$, so ist

$d(\mu, \nu) = \sup_{A \in \mathcal{K}} |\mu(A) - \nu(A)| = \frac{1}{2} \int_X |f_\mu - f_\nu| d1 = \frac{1}{2} d_1(f_\mu, f_\nu)$, wobei d_1 die Metrik des $L^1(X, \mathcal{K}, 1)$ bezeichne. (/12/, (19.36)).

M_1 ist wieder vollständig metrisierbar. Ist $\mathcal{K}$ abzählbar erzeugt, so ist $L^1(X, \mathcal{K}, 1)$ separabel (/9/, S. 227, Cor.3), und damit auch M_1 separabel.

Die Aussage des nächsten Satzes findet man auch - allerdings ohne Beweisführung - bei Norman /18/, S.226 .

(2.9) Satz Sei $((X, \mathcal{X}), (M_1, \mathcal{M}), P, u)$ ein System mit

(a) X beliebige Menge, $\mathcal{X}$ abzählbar erzeugte σ -Algebra über X .

M_1 die Menge der Wahrscheinlichkeitsmaße auf $\mathcal{X}$ mit 1-Dichte, $\mathcal{M}$ die zur starken Topologie auf M_1 gehörenden Borelschen Mengen (1 σ -endlich).

(b) $P(\mu, A) = \mu(A)$ für $\mu \in M_1$, $A \in \mathcal{X}$

(c) $u(\mu, x) = \alpha_x \mu + (1 - \alpha_x) \bigwedge_x$ für $\mu \in M_1$, $x \in X$,
wobei α_x eine $\mathcal{X}$ -meßbare Funktion von X nach $[0, 1]$
und $\bigwedge_x(A)$ ein Markovkern von $(X, \mathcal{X})$ nach $(X, \mathcal{X})$ mit
 $\bigwedge_x \in M_1$ ist.

Dann ist (X, M_1, P, u) eine lineare OM-Kette.

Beweis Gezeigt werden muß, daß die Meßbarkeitseigenschaften (1.1) b und (1.1) c erfüllt sind. (1.1) b ist mit Lemma (2.5) erledigt. Zum Nachweis von (1.1) c zeigen wir zuerst die $\mathcal{X}$ - $\mathcal{M}$ -Meßbarkeit der Abbildung $x \longrightarrow \bigwedge_x$.

Die Mengen $\{ \nu \in M_1 : d(\nu, \mu) < \varepsilon \}$, $\mu \in M_1$, $\varepsilon > 0$, erzeugen die starke Topologie von M_1 und aufgrund der Separabilität dieser Topologie (vgl. (2.8)) auch die σ -Algebra $\mathcal{M}$. Ist $\mathcal{M}$ eine abzählbare Algebra, die $\mathcal{X}$ erzeugt, so gilt unter Beachtung des folgenden Lemmas (2.10) für $\mu \in M_1$, $\varepsilon > 0$

$$\begin{aligned} & \{ x \in X : d(\bigwedge_x, \mu) < \varepsilon \} \\ &= \{ x \in X : \sup_{A \in \mathcal{X}} | \bigwedge_x(A) - \mu(A) | < \varepsilon \} \\ &= \{ x \in X : \sup_{A \in \mathcal{M}} | \bigwedge_x(A) - \mu(A) | < \varepsilon \} \\ (2.10) \quad &= \bigcap_{A \in \mathcal{M}} \{ x \in X : | \bigwedge_x(A) - \mu(A) | < \varepsilon \} \in \mathcal{X} \end{aligned}$$

da die Abbildung $x \longrightarrow \bigwedge_x(A)$ für jedes $A \in \mathcal{X}$ $\mathcal{X}$ -meßbar ist.

Gemäß Voraussetzung c) ist

$$u(\mu, x) = \alpha_x \mu + (1 - \alpha_x) \bigwedge_x$$

mit $\mathfrak{K}$ -meßbaren Funktionen α_x und $1 - \alpha_x$. Da $\bigwedge_x \mathfrak{E} - \mathcal{M}$ -meßbar ist, und weil die Addition und skalare Multiplikation stetige Abbildungen bezüglich der starken Topologie sind, ist $u(\mu, x) \mathcal{M} \times \mathfrak{K} - \mathcal{M}$ -meßbar.

(2.10) Lemma Die $\mathbb{C}$ -Algebra $\mathfrak{K}$ sei durch die abzählbare Algebra $\mathcal{M}$ erzeugt. Dann gilt für $\mu, \nu \in \mathcal{M}$

$$\sup_{A \in \mathfrak{K}} |\mu(A) - \nu(A)| = \sup_{A \in \mathcal{M}} |\mu(A) - \nu(A)|$$

Beweis Für ein Teilsystem $\tilde{\mathcal{M}} \subset \mathfrak{K}$ gelte

$$|\mu(A) - \nu(A)| \leq C \text{ für alle } A \in \tilde{\mathcal{M}}.$$

Dann gilt dies auch für alle $B = \bigcup_{i=1}^{\infty} A_i$, $A_i \in \tilde{\mathcal{M}}$, $A_i \subset A_{i+1}$,

was man so zeigt: Zu $\varepsilon > 0$ existiert aufgrund der Stetigkeit des Maßes

$$\text{ein } n_1 \in \mathbb{N} \text{ mit } |\mu(B) - \mu(A_{n_1})| \leq \varepsilon \text{ und}$$

$$\text{ein } n_2 \in \mathbb{N} \text{ mit } |\nu(B) - \nu(A_{n_2})| \leq \varepsilon$$

Für $n = \max(n_1, n_2)$ gilt dann

$$\begin{aligned} |\mu(B) - \nu(B)| &\leq \\ &\leq |\mu(B) - \mu(A_n)| + |\mu(A_n) - \nu(A_n)| + |\nu(A_n) - \nu(B)| \\ &\leq C + 2\varepsilon. \end{aligned}$$

Deswegen gilt nach einem bekannten Satz über "monotone Klassen"

(/9/, S. 12, Prop. 17) :

Aus $|\mu(A) - \nu(A)| \leq C$ für alle $A \in \mathcal{M}$ folgt $|\mu(A) - \nu(A)| \leq C$ für alle $A \in \mathfrak{K}$, womit das Lemma bewiesen ist.

(2.11) Lemma Für $x^n \in X^n$, $\mu \in S$ ist bei einer linearen OM-Kette

$$u(\mu, x^n) = \alpha_{x_n} \dots \alpha_{x_1} \mu + \alpha_{x_n} \dots \alpha_{x_2} (1 - \alpha_{x_1}) \wedge_{x_1} + \dots + (1 - \alpha_{x_n}) \wedge_{x_n}$$

Beweis Berechnet sich aus der Definition von $u(\mu, x^n)$ in

(1.1) c und (2.6) .

(2.12) Sonderfälle Für lineare OM-Ketten gilt in der Notation von Satz (1.4)

$$\begin{aligned} P_\mu(\xi_{n+1} \in A \mid \xi_1 = x_1, \dots, \xi_n = x_n) \\ &= P(u(\mu, x^n), A) = P(u(u(\mu, x^{n-1}), x_n), A) \\ &= P(\alpha_{x_n} u(\mu, x^{n-1}) + (1 - \alpha_{x_n}) \wedge_{x_n}, A) \\ &= \alpha_{x_n} P(u(\mu, x^{n-1}), A) + (1 - \alpha_{x_n}) \wedge_{x_n} (A) \\ &= \alpha_{x_n} P_\mu(\xi_n \in A \mid \xi_1 = x_1, \dots, \xi_{n-1} = x_{n-1}) + (1 - \alpha_{x_n}) \wedge_{x_n} (A) \\ \underline{\alpha_x = 0} : \end{aligned}$$

$P_\mu(\xi_{n+1} \in A \mid \xi_1 = x_1, \dots, \xi_{n-1} = x_{n-1}, \xi_n = x_n) = \wedge_{x_n} (A)$
Also bildet eine lineare OM-Kette mit $\alpha_x = 0$ für alle $x \in X$ und mit Anfangszustand μ einen Markovprozeß mit Übergangskern $\wedge_x (A)$ und mit Startverteilung μ .

$\alpha_x = 1$:

$$\begin{aligned} P_\mu(\xi_{n+1} \in A \mid \xi_1 = x_1, \dots, \xi_{n-1} = x_{n-1}, \xi_n = x_n) \\ &= P_\mu(\xi_n \in A \mid \xi_1 = x_1, \dots, \xi_{n-1} = x_{n-1}) . \end{aligned}$$

Also bildet eine lineare OM-Kette mit $\alpha_x = 1$ für alle $x \in X$ und mit Anfangszustand μ eine Kette unabhängiger, μ -verteilter Zufallsvariablen.

(2.13) Bemerkung

Lineare OM-Ketten mit beliebigem Ereignisraum X behandelt auch Bert /4/, wobei noch zusätzliche Aspekte der psychologischen Lerntheorie (in Form einer "règle d'expérimentation" nach Art der Bemerkung (4.6), 3, weiter unten) eingehen. Den Anschluß an die Theorie der Ketten mit vollständigen Bindungen stellt Bert aber nicht über den Zustandsraum $S = M/\mathfrak{X}$ her, sondern über $S = \bigcup_n X^n$ wie in Satz (1.5). Ein solches S ist aber für unsere Untersuchungen nicht brauchbar.

3. Gleichmäßig ergodische Ketten mvB

Die folgenden Bedingungen an eine Kette mvB (X, S, P, u) wurden zuerst von Norman /12/ für endliches X eingeführt und von Iosifescu und Theodorescu /13/, S. 112, auf beliebiges X ausgedehnt. Sie sind in diesem Paragraphen über die gleichmäßige Ergodizität ebenso nützlich wie in den späteren Paragraphen über Parameterschätzung.

(B_1) S ist ein kompakter metrischer Raum, $\mathfrak{S}$ das System seiner Borelschen Mengen.

(S ist dann auch vollständig und separabel)

(B_2) Für alle $A \in \mathfrak{A}$ und für alle $s, s' \in S$ gilt mit einem $B < \infty$

$$|P(s, A) - P(s', A)| \leq B d(s, s')$$

(B_3) Für alle $x \in X$ und für alle $s, s' \in S$ gilt

$$d(u(s, x), u(s', x)) \leq d(s, s')$$

(B_4) Es existieren $n_0 \in \mathbb{N}, \gamma > 0, \lambda < 1$, so daß für

$$A^{n_0} := \{ x^{n_0} \in X^{n_0} : d(u(s, x^{n_0}), u(s', x^{n_0})) \leq \lambda d(s, s') \text{ für alle } s, s' \}$$

gilt: $P(s, A^{n_0}) \geq \gamma$ für alle $s \in S$.

(A^{n_0} ist aus $\mathfrak{A}^{n_0}$, weil $d(s, s')$ und wegen (B_3) auch $d(u(s, x^{n_0}), u(s', x^{n_0}))$ stetig auf $S \times S$ sind und weil S separabel ist).

Die nächsten beiden Lemmata zeigen, unter welchen Voraussetzungen eine lineare OM-Kette die Bedingungen $(B_1) - (B_4)$ erfüllt. Wir greifen dabei auf die Notation von (2.6) und (2.8) zurück.

Zunächst formulieren wir für eine Teilmenge $S \subset M_1$ eine Bedingung:

(L) Sei $(X, \mathfrak{A}) = (\mathbb{R}^m, \mathfrak{L}^m)$, ι Lebesgue-Maß auf $\mathfrak{L}^m$ ($m \in \mathbb{N}$).

Die Konvergenzen (die für jedes $\mu \in M_\iota$ stattfinden)

$$\lim_{|y| \rightarrow 0} \int_X |f_\mu(x+y) - f_\mu(x)| \iota(dx) = 0$$

$$\lim_{n \rightarrow \infty} \int_{|x| \geq n} f_\mu(x) \iota(dx) = 0$$

sind gleichmäßig in $\mu \in S$.

(3.1) Lemma Sei (X, M_ι, P, u) eine lineare OM-Kette mit $(X, \mathfrak{A}) = (\mathbb{R}^m, \mathfrak{L}^m)$, ι Lebesgue-Maß auf $\mathfrak{L}^m$ ($m \in \mathbb{N}$). $\tilde{S}$ sei eine Teilmenge von M_ι , die Bedingung (L) erfüllt. Sei außerdem $\Lambda_x \in \tilde{S}$ für alle $x \in X$. Dann läßt sich für die Kette ein kompakter Zustandsraum S mit $\tilde{S} \subset S \subset M_\iota$ angeben (d.h.: (X, S, P, u) erfüllt (B_1)).

Beweis Setze $S_0 := \{ u(\mu, x^n) : \mu \in \tilde{S}, n \geq 0, x^n \in X^n \}$
 $(u(\mu, x^0) = \mu)$

Mit Hilfe von (2.11) rechnet man nach, daß S_0 ebenfalls

(L) erfüllt. Damit liefert ein Satz von Fréchet - Kolmogorov

(/23/, S. 275), daß S_0 präkompakt ist, die abgeschlossene

Hülle S von S_0 also kompakt ist. (Man beachte dabei die in

(2.8) erwähnte Homöomorphie von M_1 und $L^1 \cap \{ f \geq 0, \int f d\iota = 1 \}$.

Zu zeigen ist noch, daß $u(\mu, x) \in S$ für alle $\mu \in S$ gilt.

Sei also $\mu \in S$. Für jedes $\varepsilon > 0$ existiert ein $\mu_0 \in S_0$ mit

$d(\mu, \mu_0) < \varepsilon$. Dann gilt auch aufgrund von Lemma (3.2)

$$d(u(\mu, x), u(\mu_0, x)) \leq d(\mu, \mu_0) < \varepsilon.$$

Wegen $u(\mu_0, x) \in S_0$ folgt $u(\mu, x) \in S$.

(3.2) Lemma Eine lineare OM-Kette (X, S, P, u) erfüllt (B_2) ,

(B_3) . Sei $\alpha_x \mathfrak{A}$ -meßbar vorausgesetzt:

Die Bedingung $\alpha_x < 1$ für alle $x \in X$ ist notwendig für (B_4) ,

falls $S = M$, und hinreichend, falls S kompakt. In jedem Fall

folgt (B_4) aus $\alpha_x \leq \lambda$ für alle $x \in X$ ($\lambda < 1$).

Beweis

Zu (B_2) : Es ist für jedes $A \in \mathcal{A}$ und $\mu, \nu \in S$

$$|P(\mu, A) - P(\nu, A)| = |\mu(A) - \nu(A)|$$

$$\leq \sup_{A \in \mathcal{A}} |\mu(A) - \nu(A)| = d(\mu, \nu)$$

Zu (B_3) : Es gilt für jedes $x \in X$ und $\mu, \nu \in S$

$$d(u(\mu, x), u(\nu, x))$$

$$= \sup_{A \in \mathcal{A}} |\alpha_x \mu(A) + (1 - \alpha_x) \bigwedge_x(A) - \alpha_x \nu(A) - (1 - \alpha_x) \bigwedge_x(A)|$$

$$= \sup_{A \in \mathcal{A}} |\alpha_x \mu(A) - \alpha_x \nu(A)| = \alpha_x d(\mu, \nu) \leq d(\mu, \nu)$$

Zu (B_4) "notwendig": Sei angenommen, daß $A_1 := \{x \in X : \alpha_x = 1\}$

$\in \mathcal{A}$ nicht leer ist. Für $\mu \in S$ und $x_1 \in A_1$ ist $u(\mu, x_1) = \mu$.

Sei $\mu_1 \in S = M$ so gewählt, daß $\mu_1(A_1) = 1$. Wegen

$P(\mu_1, (\underbrace{A_1, \dots, A_1}_n)) = 1$ gilt dann für jedes $A^n \in \mathcal{A}^n$ mit

$P(\mu_1, A^n) > 0$, daß $A^n \cap (\underbrace{A_1, \dots, A_1}_n)$ nicht leer ist. Damit

ist (B_4) verletzt, denn für $x^n \in (\underbrace{A_1, \dots, A_1}_n)$ ist $u(\mu, x^n) = \mu$ für alle $\mu \in S$, also $d(u(\mu, x^n), u(\nu, x^n)) = d(\mu, \nu)$.

"hinreichend": Sei nun $\alpha_x < 1$ für alle $x \in X$. Setze für jedes

$N \in \mathbb{N}$ $A_N := \{x \in X : \alpha_x \leq 1 - \frac{1}{N}\}$; A_N ist mit wachsendem N

nicht abnehmend. Wegen $\bigcup_{N \in \mathbb{N}} A_N = X$ und wegen der Stetigkeit des Maßes gibt es für jedes $\mu \in S$ ein $N(\mu) \in \mathbb{N}$ mit

$P(\mu, A_{N(\mu)}) > \gamma$ ($\gamma > 0$). Da S kompakt und $P(\mu, A)$ gemäß (B_2)

für jedes $A \in \mathcal{A}$ stetig auf S ist, folgt die Existenz eines

$N \in \mathbb{N}$ mit $P(\mu, A_N) \geq \gamma$ für alle $\mu \in S$. Setze $\lambda := 1 - \frac{1}{N}$.

Wegen der oben gefundenen Beziehung

$d(u(\mu, x), u(\mu', x)) = \alpha_x d(\mu, \mu')$ ist dann (B_4) erfüllt.

(3.3) Bemerkung Ähnlich wie im eben geführten Beweis kann man zeigen, daß man (B_4) , falls $(B_1) - (B_3)$ gelten, ersetzen kann durch

$(B_4)^*$ Für alle $s \in S$ existiert ein $n = n(s)$, so daß
 $P(s, A^n) > 0$ gilt mit
 $A^n := \{ x^n \in X^n : d(u(s, x), u(s', x)) < d(s, s') \}$
 für alle $s, s' \in S$ } .

Wir kommen nun zum Begriff "gleichmäßig ergodisch", der in den Voraussetzungen der meisten folgenden Sätze zentral sein wird.

(3.4) Definition (vgl. /13/, S. 71; dort steht "ergodisch" für unser "mischend".)

Eine Kette mvB heißt gleichmäßig halb-ergodisch, falls eine Nullfolge $\varepsilon_n \geq 0$ und für jedes $l \in \mathbb{N}$ ein Markovkern $P^\infty(s, A^l)$ von $(S, \mathcal{F})$ nach $(X^l, \mathcal{A}^l)$ existieren, so daß für jedes $n \in \mathbb{N}$ gilt:

$$(*) \quad \left| \frac{1}{n} \sum_{k=1}^n P^k(s, A^l) - P^\infty(s, A^l) \right| \leq \varepsilon_n$$

für alle $s \in S$, $l \in \mathbb{N}$, $A^l \in \mathcal{A}^l$.

Gilt $(*)$ sogar mit $P^n(s, A^l)$ anstelle von $\frac{1}{n} \sum_{k=1}^n P^k(s, A^l)$, so nennen wir die Kette gleichmäßig halb-mischend.

(3.5) Definition Ist in (3.4) für jedes $l \in \mathbb{N}$, $A^l \in \mathcal{A}^l$ $P^\infty(s, A^l) = P^\infty(A^l)$ unabhängig von s , so nennen wir die Kette mvB gleichmäßig ergodisch bzw. gleichmäßig mischend.

(3.6) Satz Sind für eine Kette mvB $(B_1) - (B_4)$ erfüllt, so ist sie gleichmäßig halb-ergodisch.

Darüberhinaus ist $P^\infty(s, A^l)$ für jedes $l \in \mathbb{N}$, $A^l \in \mathcal{A}^l$, stetig in s .

Beweis /17/, Theorem 2.1 und Lemma 6.1 für endliches X und
/13/, Theorem 2.1.64 für beliebiges X .

(3.7) Definition Für jedes $s \in S$, $n \in \mathbb{N}$ setzen wir

$$T_n(s) := \bigcap \{ B \in \mathcal{O} : B \text{ abgeschlossen und } Q^n(s, B) = 1 \}$$
 $(Q^n \text{ siehe (1.10)})$. $T_n(s)$ ist abgeschlossen, und zufolge der
 abzählbaren Basis von S gilt $Q^n(s, T_n(s)) = 1$.

Die folgende Bedingung wurde bereits von Jamison, /14/,
 Theorem 3.4, benutzt, um die gleichmäßige Ergodizität zu
 beweisen.

$(B_5)^*$ Für alle $s, s' \in S$ ist $\lim_n d(T_n(s), T_n(s')) = 0$

(3.8) Satz Gilt für eine Kette mvB $(B_1) - (B_4)$, $(B_5)^*$, so
 ist sie gleichmäßig mischend.

Beweis /17/, Theorem 2.4 für endliches X und
 /13/, Theorem 2.1.65 für beliebiges X .

Wir wollen nun die Bedingung $(B_5)^*$ so abschwächen, daß
 "periodische" Prozesse nicht wie bei $(B_5)^*$ ausgeschlossen
 werden.

(B_5) Für alle $s, s' \in S$ ist $d(\bigcup_{n \in \mathbb{N}} T_n(s), \bigcup_{n \in \mathbb{N}} T_n(s')) = 0$

(3.9) Satz Gilt für eine Kette mvB $(B_1) - (B_5)$, so ist sie
 gleichmäßig ergodisch.

Beweis Die nun vorgeführte Beweisidee geht auf Jamison /14/
 zurück und wurde auch von Norman im Beweis des Satzes (3.8)
 benutzt: Als erstes zeigen wir, daß die Funktion

$$\phi(s) := P^\infty(s, A^1) = \lim_n \frac{1}{n} \sum_{k=1}^n P^k(s, A^1)$$

aus Satz (3.6) ein Fixpunkt der Abbildung

$$\phi \longrightarrow \int_S Q(\cdot, ds') \phi(s') \text{ ist.}$$

$$\int_S Q(s, ds') \phi(s')$$

$$= \int_X P(s, dx) P^\infty(u(s, x), A^1)$$

$$= \int_X P(s, dx) \lim_n \frac{1}{n} \sum_{k=1}^n P^k(u(s, x), A^1)$$

$$= \lim_n \frac{1}{n} \sum_{k=1}^n \int_X P(s, dx) P^k(u(s, x), A^1) \quad (\text{Satz von der majorisierten Konvergenz})$$

$$= \lim_n \frac{1}{n} \sum_{k=1}^n P^{k+1}(s, A^1)$$

$$= P^\infty(s, A^1) = \phi(s)$$

Es folgt dann auch für alle $n \in \mathbb{N}$

$$\int_S Q^n(s, ds') \phi(s') = \phi(s)$$

Wir müssen zeigen, daß $\phi(s)$ eine Konstante ist. $\phi(s)$ ist

nach Satz (3.6) stetig auf dem Kompaktum S : Deshalb setzt man

$$M = \max_s \phi(s) = \phi(s_M), \quad m = \min_s \phi(s) = \phi(s_m), \quad s_M, s_m \in S$$

$$B_M := \{s \in S : \phi(s) = M\}, \quad B_m := \{s \in S : \phi(s) = m\}$$

B_M und B_m sind abgeschlossen.

Aus $\int_S Q^n(s_M, ds') \phi(s') = \phi(s_M)$ und $\phi(s_M) \geq \phi(s)$ für $s \in S$ folgt

$$Q^n(s_M, B_m) = 1 \text{ für alle } n \in \mathbb{N}, \text{ und völlig analog auch:}$$

$$Q^n(s_m, B_M) = 1 \text{ für alle } n \in \mathbb{N}.$$

Also gilt $T_n(s_M) \subset B_m$ für alle $n \in \mathbb{N}$ und $T_n(s_m) \subset B_M$

für alle $n \in \mathbb{N}$.

Mache nun die Annahme: $M - m > 0$:

Aus $|\phi(s') - \phi(s'')| = M - m > 0$ für $s' \in B_m, s'' \in B_M$ folgt

wegen der Stetigkeit von $\phi(s)$, daß es ein $\delta > 0$ gibt mit

$d(B_m, B_M) \geq \delta$ und daraus

$$d(\bigcup_n T_n(s_m), \bigcup_n T_n(s_M)) \geq \delta,$$

was im Widerspruch zu (B_5) steht.

Für den Rest des Kapitels wollen wir uns nicht mehr mit den Voraussetzungen (B_1) beschäftigen, sondern gleich voraussetzen, daß die Kette mvB gleichmäßig ergodisch ist.

Für die miteinander verträglichen Maße $P^\infty|_{\mathcal{X}^1}$, $1 \in \mathbb{N}$, kann ohne zusätzliche topologische Voraussetzungen die Existenz eines Maßes P auf $\mathcal{X}^{\mathbb{N}}$ gesichert werden, das auf den $A^1 \in \mathcal{X}^1$ die Werte $P^\infty(A^1)$ annimmt.

(3.10) Satz Es ex. ein Wahrscheinlichkeitsmaß P auf $\mathcal{X}^{\mathbb{N}}$, so daß für alle $1 \in \mathbb{N}$, $A^1 \in \mathcal{X}^1$ gilt:

$$P(A^1) = P^\infty(A^1)$$

Beweis (nach Theiler /25/) Wir haben für jedes $n \in \mathbb{N}$

$$(*) \quad \left| \frac{1}{n} \sum_{k=1}^n P^k(s, A^1) - P^\infty(A^1) \right| \leq \varepsilon_n$$

für alle $1 \in \mathbb{N}$, $A^1 \in \mathcal{X}^1$, $s \in S$.

Analog zu /25/, S. 196, zeigt man, daß es für jedes $A^{\mathbb{N}} \in \mathcal{X}^{\mathbb{N}}$ eine Zahl $P^\infty(A^{\mathbb{N}})$ gibt, so daß $(*)$ mit $A^{\mathbb{N}}$ anstelle von A^1 gilt.

(Man beachte, daß $P^k(s, A^{\mathbb{N}})$ gemäß (1.7) eine Wahrscheinlichkeit auf $\mathcal{X}^{\mathbb{N}}$ ist.) Eine Ausführung des Beweises findet man im

ANHANG. Die Gleichmäßigkeit der Konvergenz in $(*)$ garantiert, daß

$$(**) \quad P^\infty(A^{\mathbb{N}}) = \lim_n \frac{1}{n} \sum_{k=1}^n P^k(s, A^{\mathbb{N}})$$

ein Wahrscheinlichkeitsmaß auf $\mathcal{X}^{\mathbb{N}}$ ist.

Dieses Wahrscheinlichkeitsmaß, das auf den $A^1 \in \mathcal{X}^1$ mit $P^\infty(A^1)$ übereinstimmt, soll mit $P|_{\mathcal{X}^{\mathbb{N}}}$ bezeichnet werden.

(3.11) Definition Die Abbildung $T : X^{\mathbb{N}} \longrightarrow X^{\mathbb{N}}$

(bzw: $X^{\mathbb{Z}} \longrightarrow X^{\mathbb{Z}}$) mit

$(Tx^{\mathbb{N}})_n = x_{n+1}$ für alle $n \in \mathbb{N}$ (bzw: $(Tx^{\mathbb{Z}})_n = x_{n+1}$ für alle $n \in \mathbb{Z}$)

heißt Verschiebungsoperator.

(3.12) Satz Der Operator T ist maßerhaltend bezüglich

$P|_{\mathcal{X}^{\mathbb{N}}}$. (Wir nennen P dann auch stationäres Wahrscheinlichkeitsmaß)

Beweis Nach /6/, Theorem 1.1, reicht es zu zeigen, daß

$P(A^1) = P(T^{-1}A^1)$ für alle $1 \in \mathbb{N}$, $A^1 \in \mathcal{X}^1$, oder,

was dasselbe ist,

$P(A^1) = P((X, A^1))$ mit $(X, A^1) \in \mathcal{X}^{1+1}$. Nun ist

$$\begin{aligned} P(A^1) &= \lim_n \frac{1}{n} \sum_{k=1}^n P^k(s, A^1) = \lim_n \frac{1}{n} \sum_{k=1}^n P^{k+1}(s, A^1) \\ &= \lim_n \frac{1}{n} \sum_{k=1}^n P^k(s, (X, A^1)) = P((X, A^1)) \end{aligned}$$

(3.13) Definition Der Verschiebungsoperator T heißt

ergodisch (oder auch: metrisch transitiv) bezüglich der

stationären Wahrscheinlichkeit $P|_{\mathcal{X}^{\mathbb{N}}}$, wenn gilt:

$T^{-1}A^{\mathbb{N}} = A^{\mathbb{N}}$ für ein $A^{\mathbb{N}} \in \mathcal{X}^{\mathbb{N}}$ $\implies P(A^{\mathbb{N}}) = 0$ oder 1

(3.14) Satz Der Operator T ist ergodisch bezüglich $P|_{\mathcal{X}^{\mathbb{N}}}$.

Beweis Gemäß /6/, Theorem 1.4, reicht es aus, wenn für alle

$A^1 \in \mathcal{X}^1$ und $A^k \in \mathcal{X}^k$ gilt:

$$\lim_n \frac{1}{n} \sum_{i=1}^n P(A^1 \cap T^{-i}A^k) = P(A^1) \cdot P(A^k)$$

Nun ist für beliebiges $s \in S$

$$\frac{1}{n} \sum_{i=1}^n P(A^1 \cap T^{-i}A^k)$$

$$= \frac{1}{n} \sum_{i=1}^n \lim_m \frac{1}{m} \sum_{j=1}^m P(s, T^{-j}A^1 \cap T^{-(j+i)}A^k)$$

Nach (1.8) ist, sofern $i \geq 1$ ist,

$$\begin{aligned} & P(s, T^{-j} A^1 \cap T^{-(j+i)} A^k) \\ &= \int_{X A^1} P(s, dx^{j+1}) P(u(s, x^{j+1}), T^{-(i-1)} A^k) \end{aligned}$$

Damit ist

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n P(A^1 \cap T^{-i} A^k) \\ &= \lim_m \frac{1}{m} \sum_{j=1}^m \int_{X A^1} P(s, dx^{j+1}) \frac{1}{n} \sum_{i=1}^n P(u(s, x^{j+1}), T^{-(i-1)} A^k) \\ &= \lim_m \frac{1}{m} \sum_{j=1}^m \int_{X A^1} P(s, dx^{j+1}) \cdot P(A^k) + \varepsilon_n^*, \text{ wobei} \\ & |\varepsilon_n^*| \leq \sup_{s'} \left| \frac{1}{n} \sum_{i=1}^n P(s', T^{-(i-1)} A^k) - P(A^k) \right| \end{aligned}$$

Wegen der gleichmäßigen Ergodizität gilt die Ungleichung $|\varepsilon_n^*| \leq \varepsilon_n$, wobei ε_n eine Nullfolge ist.

Damit haben wir

$$\begin{aligned} & \lim_n \frac{1}{n} \sum_{i=1}^n P(A^1 \cap T^{-i} A^k) \\ &= \lim_m \frac{1}{m} \sum_{j=1}^m \int_{X A^1} P(s, dx^{j+1}) \cdot P(A^k) \\ &= \lim_m \frac{1}{m} \sum_{j=1}^m P^j(s, A^1) \cdot P(A^k) \\ &= P(A^1) \cdot P(A^k) \quad \text{w.z.b.w.} \end{aligned}$$

(3.15) Bemerkung Ist der Operator T ergodisch bezüglich der stationären Wahrscheinlichkeit $P|_{\mathcal{B}^{\mathbb{N}}}$, und ist $f(x^{\mathbb{N}})$ eine meßbare Funktion auf $X^{\mathbb{N}}$ mit

$$E|f| = \int |f(x^{\mathbb{N}})| \cdot P(dx^{\mathbb{N}}) < \infty,$$

dann gilt das "individuelle Ergodentheorem" (/6/, Theorem 1.3)

$$\lim_n \frac{1}{n} \sum_{k=1}^n f(T^k x^{\mathbb{N}}) = E f \quad \text{für } P\text{-fast alle } x^{\mathbb{N}}.$$

Dieses Ergodentheorem bildet ein wichtiges Werkzeug in den späteren Untersuchungen. Dabei wird stets das Problem auf-

tauchen, daß wir die Konvergenz nicht P -fast überall, sondern P_S -fast überall benötigen (P_S ist in (1.7) definiert). Denn P_S ist diejenige Wahrscheinlichkeit, die einer Realisierung der Kette mvB von einem Startzustand s aus tatsächlich zugrunde liegt. Dieses Problem gehen wir mit (3.16), (3.17) an. Im folgenden Satz werden $\mathcal{K}^{\mathbb{N}}$ und $\mathcal{K}^{[n, \infty)}$, $n \in \mathbb{N}$, als Unter- $\mathcal{G}$ -Algebren von $\mathcal{K}^{\mathbb{Z}}$ betrachtet. Die Bezeichnung $A \in \mathcal{K}^{\mathbb{N}}$ soll dann über die Meßbarkeit von A hinaus betonen, daß $P(s, A)$ auch wirklich definiert ist.

(3.16) Definition (vgl. /1/, Def. 30.6)

$\bigcap_{n \in \mathbb{N}} \mathcal{K}^{[n, \infty)}$ heißt die $\mathcal{G}$ -Algebra der terminalen Ereignisse aus $\mathcal{K}^{\mathbb{N}}$.

(3.17) Satz Sei $A \in \mathcal{K}^{\mathbb{N}}$ ein terminales Ereignis mit $P(A) = 0$. Dann ist auch $P(s, A) = 0$ für alle $s \in S$.

Beweis Für ein terminales A ist $T^n A \in \mathcal{K}^{\mathbb{N}}$ für jedes $n \in \mathbb{N}$. Denn es ist $A \in \mathcal{K}^{[n+1, \infty)}$, woraus $T^n A \in \mathcal{K}^{\mathbb{N}}$ folgt. Also ist auch $B := \bigcup_{n=-\infty}^{\infty} T^n A$ aus $\mathcal{K}^{\mathbb{N}}$, ferner $B \supset A$.

Aus $P(A) = 0$ folgt nun $P(B) = 0$ wegen

$$P(B) = P\left(\bigcup_{n=-\infty}^{\infty} T^n A\right) \leq \sum_{n=-\infty}^{\infty} P(T^n A) = \sum_{n=-\infty}^{\infty} P(A) = 0.$$

Es gilt $T^{-k} B = \bigcup_{n=-\infty}^{\infty} T^{n-k} A = \bigcup_{n=-\infty}^{\infty} T^n A = B$ und deshalb:

$$0 = P(A) = P(B) \stackrel{(3.10) (**)}{=} \lim_n \frac{1}{n} \sum_{k=1}^n P(s, T^{-k} B)$$

$$= \lim_n \frac{1}{n} \sum_{k=1}^n P(s, B) = P(s, B) \geq P(s, A) \text{ für alle } s \in S \quad \text{w.z.b.w.}$$

In /13/, Prop. 2.2.10, wird bewiesen, daß für gleichmäßig mischende Ketten mvB sogar $P(A) = P(s, A)$ für alle terminalen Ereignisse $A \in \mathfrak{E}^{\mathbb{N}}$ gilt. (3.17) zeigt, was sich von diesem Ergebnis bei gleichmäßiger Ergodizität noch retten läßt. Für spätere Zwecke brauchen wir noch die beiden folgenden Sätze.

(3.18) Satz Für ein $A \in \mathfrak{E}$ und ein $\gamma > 0$ gelte

$P(s, A) \geq \gamma$ für alle $s \in S$.

Setze $\nu^n(x^{\mathbb{N}}) := \sum_{i=1}^n \chi_A(x_i)$.

Dann existiert ein $\alpha > 0$ und für P-fast alle $x^{\mathbb{N}}$ ein $N = N(x^{\mathbb{N}})$ mit

$\nu^n(x^{\mathbb{N}}) \geq \alpha n$ für alle $n \geq N$

Beweis Das Ergodentheorem (3.15) liefert für P-fast alle $x^{\mathbb{N}}$:

$$\frac{1}{n} \nu^n(x^{\mathbb{N}}) = \frac{1}{n} \sum_{i=1}^n \chi_A(x_i) \longrightarrow E \chi_A = P(A).$$

$$\text{Nun ist } P(A) = \lim_n \frac{1}{n} \sum_{k=1}^n P(s, x^k A)$$

$$= \lim_n \frac{1}{n} \sum_{k=1}^n \int_{X^k} P(s, dx^k) P(u(s, x^k), A)$$

$$\geq \gamma \lim_n \frac{1}{n} \sum_{k=1}^n P(s, x^k) = \gamma P(x^k) = \gamma.$$

Setze $\alpha := P(A)/2 \geq \gamma/2 > 0$.

Es existiert dann ein $N = N(x^{\mathbb{N}})$ mit

$$\frac{1}{n} \nu^n(x^{\mathbb{N}}) \geq \alpha \text{ für alle } n \geq N. \quad \text{w.z.b.w.}$$

(3.19) Satz Für ein $A \in \mathfrak{E}$ gelte $P(A) < 1$.

Dann existieren positive Zahlen $c < \infty$, $d < 1$ mit

$$P(\underbrace{(A, \dots, A)}_n) \leq cd^n \text{ für alle } n \in \mathbb{N}.$$

Beweis Sei $P(A) < \delta < 1$. Wegen der gleichmäßigen Ergodizität existiert ein $n_0 \in \mathbb{N}$, so daß

$$\frac{1}{n_0} \sum_{k=1}^{n_0} P^k(s, A) \leq \delta \text{ für alle } s \in S.$$

Das aber heißt: Es gibt für jedes $s \in S$ ein $k(s)$,

$$1 \leq k(s) \leq n_0, \text{ mit } P^{k(s)}(s, A) \leq \delta.$$

Daraus folgt, wenn wir innerhalb dieses Beweises verabreden,

daß $A^n = \underbrace{(A, \dots, A)}_n$ bedeutet:

$$P(s, A^{n_0}) \leq P^{k(s)}(s, A) \leq \delta \text{ für alle } s \in S.$$

Nun zeigen wir durch Induktion, daß für alle $k \in \mathbb{N}$

$$P(A^{n_0 \cdot k}) \leq \delta^k \text{ ist.}$$

$$P(A^{n_0 \cdot k}) = \lim \frac{1}{n} \sum_{i=1}^n P^{i+1}(s, A^{n_0 \cdot k})$$

$$\stackrel{(1.8)}{=} \lim \frac{1}{n} \sum_{i=1}^n \int_{X^{i \cdot n_0 \cdot (k-1)}} P(s, dx^i x^{n_0 \cdot (k-1)}).$$

$$= P(u(s, x^{i \cdot n_0 \cdot (k-1)}), A^{n_0})$$

$$\leq P(A^{n_0 \cdot (k-1)}) \cdot \delta \stackrel{(\text{Ind.vor})}{\leq} \delta^{(k-1)} \cdot \delta = \delta^k$$

Mit $c := \delta^{-1} < \infty$ und $d := \delta^{\frac{1}{n_0}} < 1$ ist nun die Behauptung bewiesen.

4. Limesverhalten endlicher linearer OM-Ketten

(4.1) Erläuterungen In Spezialisierung der Definition

(2.6) gilt:

Eine endliche lineare OM-Kette besteht aus den Elementen

$$X := \{1, 2, \dots, N\}, \quad N \in \mathbb{N}, \quad X = \mathcal{P}(X)$$

$S := \{p \in \mathbb{R}^N : p_i \geq 0, \sum_{i=1}^N p_i = 1\}$, $\mathcal{O}$ = Borelsche Teilmengen von S ($p \in \mathbb{R}^N$ wird in diesem Paragraphen stets als Zeilenvektor aufgefaßt.)

$$P(p, i) = p_i \quad \text{für alle } p \in S, i \in X$$

$$u(p, i) = \alpha_i p + (1 - \alpha_i) \bigwedge_i,$$

$$\text{wobei } \alpha = (\alpha_1, \dots, \alpha_N) \in \mathbb{R}^N, \quad 0 \leq \alpha_i \leq 1,$$

$\bigwedge$ eine stochastische $N \times N$ Matrix ist und $\bigwedge_i$ ihre i -te Zeile.

Eine endliche lineare OM-Kette wird also durch Angabe der Größen $(\alpha, \bigwedge)$ spezifiziert.

(4.2) Lemma Eine endliche lineare OM-Kette $(\alpha, \bigwedge)$ mit $\alpha_i < 1$ für alle $i \in X$ erfüllt die Bedingungen $(B_1) - (B_4)$.

Beweis (B_1) gilt, weil das in (4.1) eingeführte S kompakt ist.

$(B_2) - (B_4)$ gilt wegen Lemma (3.2).

Wir wollen nun zeigen, daß sich -grob gesagt- das Limesverhalten einer Kette $(\alpha, \bigwedge)$ zurückführen läßt auf das Limesverhalten von $\bigwedge$.

(4.3) Definitionen Eine nicht negative Matrix $\bigwedge$ heißt irreduzibel, falls es für alle $i, j \in X$ ein $n = n(i, j)$ gibt mit $(\bigwedge^n)_{ij} > 0$.

Eine irreduzible Matrix heißt aperiodisch, falls es ein n gibt

mit $(\wedge^n)_{ij} > 0$ für alle $i, j \in X$.

(Vgl. /10/, chap. XV und /6/, S. 119-121)

Wir wollen nun eine stochastische Matrix $\wedge$ ergodisch nennen, falls sie durch Umnummerieren der Elemente aus X auf die folgende Form gebracht werden kann

$$(*) \wedge = \begin{pmatrix} R & O \\ Q & S \end{pmatrix} \quad \begin{array}{l} R, S \text{ quadratische Matrizen, } O \text{ Nullmatrix} \\ R \text{ irreduzibel, } S^n \longrightarrow 0 \text{ für } n \longrightarrow \infty \end{array}$$

Eine stochastische Matrix $\wedge$ wollen wir mischend nennen, wenn sie ergodisch ist und das R in der Darstellung $(*)$ aperiodisch ist.

(4.4) Lemma Eine stochastische Matrix $\wedge$ sei

(a) ergodisch. Dann gilt für die Matrix $\wedge^\infty := \lim_n \frac{1}{n} \sum_{k=1}^n \wedge^k$,
(die für jede stochastische Matrix $\wedge$ existiert)

$$\wedge_{ij}^\infty = \wedge_j^\infty$$

(b) mischend. Dann existiert $\wedge^\infty = \lim_n \wedge^n$ und es gilt:

$$\wedge_{ij}^\infty = \wedge_j^\infty$$

Beweis /8/, S. 125-129 und /10/, chap. XV

Dieses Lemma rechtfertigt die Begriffsbildung in Definition

(4.3), denn für eine ergodische Matrix $\wedge$ ist die zugehörige

Markovkette gleichmäßig ergodisch im Sinne der Definition (3.5),

für eine mischende Matrix $\wedge$ gleichmäßig mischend.

(4.5) Satz Für eine endliche lineare OM-Kette $(\alpha, \wedge)$ mit

$\alpha_i < 1$ für alle $i \in X$ gelte:

(a) $\wedge$ ergodisch: Dann ist die Kette $(\alpha, \wedge)$ gleichmäßig ergo-
disch.

(b) $\wedge$ mischend: Dann ist die Kette $(\alpha, \wedge)$ gleichmäßig mischend

Beweis Mit $i^n \in X^n$, $i^n = (i_1, \dots, i_n)$ gilt für jedes $p \in S$ nach (2.11)

$$u(p, i^n) = \alpha_{i_n} \dots \alpha_{i_1} p + \alpha_{i_n} \dots \alpha_{i_2} (1 - \alpha_{i_1}) \wedge i_1 + \dots + (1 - \alpha_{i_n}) \wedge i_n$$

und deshalb mit $\alpha_{\max} := \max_i \alpha_i$, $\alpha_{\max} < 1$ und Euklidischer Metrik d

$d(u(\tilde{p}, i^n), u(\tilde{\tilde{p}}, i^n)) \leq \sqrt{2} \alpha_{\max}^n$ für alle $\tilde{p}, \tilde{\tilde{p}} \in S$. Ist ferner für ein $i^n \in X^n$ $P^m(p, i^n) > 0$, so gibt es ein $\tilde{p} \in S$ mit

$u(\tilde{p}, i^n) \in T_{m+n}(p)$. Deswegen ist (B_5) erfüllt, falls gilt:

$$(1) \begin{cases} \text{Zu } p', p'' \in S \text{ gibt es für alle } n \text{ ein } i^n \in X^n \text{ und } n', n'' \in \mathbb{N}, \\ \text{so daß } P^{n'}(p', i^n) > 0 \text{ und } P^{n''}(p'', i^n) > 0 \end{cases}$$

und $(B_5)^*$ ist erfüllt, falls gilt:

$$(1)^* \{ \text{In (1) gibt es } n', n'' \text{ mit } n' = n'' \}.$$

Für jedes $p \in S$ und $k \in X$ gilt nun

$$(*) \quad u(p, k) \geq (1 - \alpha_k) \wedge_k \quad \text{und so auch} \quad P(u(p, k), \cdot) \geq (1 - \alpha_k) \wedge_k.$$

Deswegen folgt aus

$$(2) \begin{cases} \text{Zu } p', p'' \in S \text{ gibt es ein } k \in X \text{ und } n', n'' \in \mathbb{N}, \text{ so daß} \\ P^{n'}(p', k) > 0 \text{ und } P^{n''}(p'', k) > 0 \end{cases}$$

daß es ein $i^n = (k, i_2, \dots, i_n) \in X^n$ gibt mit

$$P^{n'}(p', i^n) > 0 \text{ und } P^{n''}(p'', i^n) > 0.$$

(1) ist also erfüllt, falls (2) gilt, und $(1)^*$ ist erfüllt, falls gilt:

$$(2)^* \{ \text{in (2) gibt es } n', n'' \text{ mit } n' = n'' \}.$$

Mit der Abkürzung $A_{ij} := (1 - \alpha_i) \wedge_{ij}$ gilt für jedes

$p = (p_1, \dots, p_N) \in S$ die Ungleichung

$$P^n(p, k) \geq \sum_{i=1}^N p_i A_{ik}^{n-1}$$

Beweis durch Induktion:

Der Induktionsanfang ist wegen $P^1(p, k) = p_k$ und

$$\sum_i p_i A_{ik}^0 = \sum_i p_i \delta_{ik} = p_k \text{ gegeben.}$$

$$p^{n+1}(p, k) \stackrel{(1.8)}{=} \sum_i p_i p^n(u(p, i), k)$$

$$(\text{Ind. Vor.}) \sum_i p_i \sum_j u(p, i)_j A_{jk}^{n-1}$$

$$\stackrel{\geq}{(*)} \sum_i p_i \sum_j (1 - \alpha_i) \wedge_{ij} A_{jk}^{n-1}$$

$$= \sum_i p_i \sum_j A_{ij} A_{jk}^{n-1} = \sum_i p_i A_{ik}^n.$$

Nun können wir den Beweis beenden:

(a) Sei $\wedge$ ergodisch

Zu $p', p'' \in S$ wähle $i' \in X$, $i'' \in X$ mit $p_{i'}' > 0$ und $p_{i''}'' > 0$.

Es existiert ein k und n' , $n'' \in \mathbb{N}$ mit $\bigwedge_{i', k}^{n'} > 0$ und $\bigwedge_{i'', k}^{n''} > 0$

Wegen $(1 - \alpha_i) > 0$ für alle i folgt daraus $A_{i', k}^{n'} > 0$ und $A_{i'', k}^{n''} > 0$ und

$$p^{n'+1}(p', k) \geq p_{i'}' A_{i', k}^{n'} > 0, p^{n''+1}(p'', k) \geq p_{i''}'' A_{i'', k}^{n''} > 0$$

Damit ist (2) erfüllt, also auch (1), also auch (B_5) und damit die Behauptung (a).

(b) Sei $\wedge$ mischend.

Dann kann man in (a) $n = n' = n''$ wählen und man kommt auf ein $k \in X$ mit

$$p^n(p', k) > 0 \text{ und } p^n(p'', k) > 0.$$

Damit ist $(2)^*$ erfüllt, also auch $(1)^*$, also auch $(B_5)^*$ und damit die Behauptung (b).

(4.6) Bemerkungen

1. Setzt man in Satz (4.5) alle $\alpha_i = 0$, so erhält man Lemma (4.4) zurück.

2. In Satz (4.5) zeigt sich sehr konkret die Nützlichkeit der durch Bedingung (B_5) und Satz (3.9) vorgenommenen Erweiterung gegenüber Bedingung $(B_5)^*$ und Satz (3.8), weil durch sie auch die Klasse der periodischen (=nicht aperiodischen) Matrizen $\wedge$ mit in die Betrachtung eingeschlossen wird.

3. In der psychologischen Lerntheorie betrachtet man die folgende Erweiterung der linearen OM -Kette: Eine Kette mvB heißt "lineares Lernmodell" (geauer: $N \times M$ -Operator Modell, vgl. /13/, S.225), wenn gilt:

$$X = \{(i,j) : 1 \leq i \leq N, 1 \leq j \leq M\}, N, M \in \mathbb{N}$$

S die Menge der N -dimensionalen Wahrscheinlichkeitsvektoren, und für $(i,j) \in X, p \in S$

$$P(p, (i,j)) = p_i \Pi_{ij} \quad \text{und}$$

$$u(p, (i,j)) = \alpha_{ij} p + (1 - \alpha_{ij}) \Lambda_{ij},$$

wobei (Π_{ij}) eine stochastische $N \times M$ Matrix ist, $0 \leq \alpha_{ij} \leq 1$,

und $\Lambda_{ij} \in S$ (d.h. $(\Lambda_{ij,k})$ eine stochastische $N \times M$ Matrix).

Dann läßt sich der obige Beweis auf diese Situation so erweitern, daß gilt: Ein lineares Lernmodell mit

$\alpha_{ij} < 1$ für alle $(i,j) \in X$ ist gleichmäßig ergodisch (gleichmäßig mischend), wenn die stochastische $N \times M$ Matrix

$B_{ij,kl} := \Lambda_{ij,k} \Pi_{kl}$ ergodisch (mischend) im Sinne von (4.3)

ist. Anstelle von " $\alpha_{ij} < 1$ für alle $(i,j) \in X$ " reicht auch "für

alle i gibt es ein j mit $\alpha_{ij} < 1$ und $\Pi_{ij} > 0$ " aus, wenn man die

stochastische Matrix B auf dem reduzierten Raum $\{(i,j) \in X :$

$\alpha_{ij} < 1\}$ betrachtet. Dies stellt dann eine Erweiterung eines

Satzes von Norman /18/, S.227, Th. 1.2, 1.3, dar, wenn man die-

sen auf den Fall eines endlichen Raumes X (dort E) projiziert.

TEIL II Parameterschätzung in gleichmäßig ergodischen
Ketten mvB

5. Dichten für die Übergangswahrscheinlichkeiten $P(s,A)$

In Hinblick auf das Problem der Parameterschätzung wollen wir nun annehmen, daß die Wahrscheinlichkeit $P(s,A)$, $A \in \mathfrak{A}$, für jedes $s \in S$ Dichten besitzt. Das Ziel der nächsten beiden Paragraphen ist es, für diese Dichten eine Version des Shannon-McMillan Theorems zu beweisen (vgl. Satz (6.5)). Dabei wird noch keine Abhängigkeit der zur Kette mvB (X,S,P,u) gehörenden Größen P und u von einem Parameter ϑ postuliert, weil dies die Lesbarkeit der Beweise sehr belasten würde. In Paragraph 7 lassen sich dann die Ergebnisse aus den Paragraphen 5 und 6 sehr leicht auf den Fall übertragen, daß P und u von einem ϑ abhängen. Die Voraussetzungen, die wir zur Herleitung von Satz (6.5) machen, hängen eng mit den Voraussetzungen $(B_1) - (B_4)$ zusammen. Dies ist deswegen ökonomisch, weil wir in Satz (6.5) auch die gleichmäßige Ergodizität voraussetzen müssen, die ja aus $(B_1) - (B_4)$, (B_5) folgt.

(C_1) Für eine Kette mvB (X,S,P,u) gelte:

S ist vollständig metrisierbar mit $d(s,s') \leq D$ für alle $s,s' \in S$ ($D < \infty$).

Bezüglich eines Wahrscheinlichkeitsmaßes $l|_{\mathfrak{A}}$ besitzt jedes Wahrscheinlichkeitsmaß $P(s,\cdot)|_{\mathfrak{A}}$ eine Dichte $p(s,x)$; also

$$P(s,A) = \int_A p(s,x) l(dx) \quad \text{für jedes } A \in \mathfrak{A}.$$

Für jedes $n \in \mathbb{N}$, $s \in S$ ist $p(u(s,x^n),x)$

$\mathfrak{A}^n \times \mathfrak{A}$ -meßbar.

Das nächste Lemma sagt aus, daß man sich aus einem vorgegebenem S eventuell ein S^* mit der Eigenschaft (C_1) auswählen kann.

(5.1) Lemma Für die Kette mvB (X, S, P, u) gelte (B_2) . Ist S vollständig metrisierbar, dann auch $S^* := \{s \in S: P(s, \cdot) | \mathfrak{A} \text{ ist } 1\text{-stetig}\}$. Ist S sogar polnisch, so ist bei geeignetem 1 $S = S^*$.

Beweis Sei s_n eine Cauchyfolge aus S^* (o.E. S^* nicht leer).

Für $s = \lim_n s_n$, $s \in S$, gilt wegen (B_2) $P(s, A) = \lim_n P(s_n, A)$ für alle $A \in \mathfrak{A}$, woraus die 1-Stetigkeit von $P(s, \cdot) | \mathfrak{A}$ folgt.

Ist S noch separabel, so folgt aus (B_2) , daß die Menge der W.maße $\{P(s, \cdot) | \mathfrak{A}, s \in S\}$ in der starken Topologie (Def. (2.4)) ebenfalls separabel ist. Sei $\{P(s_n, \cdot) | \mathfrak{A}, n \in \mathbb{N}\}$ die abzählbar dichte Menge. Dann ist $1(A) := \sum_n 2^{-n} P(s_n, A)$, $A \in \mathfrak{A}$, ein W.maß auf $\mathfrak{A}$, das alle $P(s_n, \cdot)$ und damit alle $P(s, \cdot)$ dominiert.

Bemerkung S^* braucht nicht gegen die Abbildung $u(\cdot, x): S \rightarrow S$ für alle $x \in X$ abgeschlossen zu sein, ist es aber im Fall der lin. OM-Kette, falls auch alle Λ_x 1-stetig sind.

(5.2) Lemma Das Maß $P(s, A^n)$, $A^n \in \mathfrak{A}^n$, $(n \in \mathbb{N})$ besitzt die Dichte

$$p(s, x^n) = \prod_{k=1}^n p(u(s, x^{k-1}), x_k)$$

bezüglich 1^n . ($u(s, x^0) := s$)

Beweis Gemäß (1.8) gilt für jedes $A^n \in \mathfrak{A}^n$

$$P(s, A^n) = \int_{A^n} p(s, x_1) p(u(s, x_1), x_2) \dots p(u(s, x^{n-1}), x_n) 1^n(dx^n),$$

was die Behauptung liefert.

Nun stellen wir an die Kette mvB (X, S, P, u) die weitere Bedingung (C_2) Für alle $s, s' \in S$, $n \in \mathbb{N}$, $x^n \in X^n$, $x \in X$ gilt mit einem $C < \infty$

$$|p(u(s, x^n), x) - p(u(s', x^n), x)| \leq C \frac{d(u(s, x^n), u(s', x^n))}{d(s, s')}$$

(wobei C noch von x abhängen kann)

(5.3) Bemerkungen Man hat sich (C_2) als eine Verallgemeinerung der zu (B_2) analogen Bedingung

$$|p(s, x) - p(s', x)| \leq C d(s, s')$$

vorzustellen. Aus dieser folgt nämlich (C_2) wie folgt:

$$\begin{aligned} & |p(u(s, x^n), x) - p(u(s', x^n), x)| \leq C d(u(s, x^n), u(s', x^n)) \\ & = \frac{(C \cdot D)}{D} d(u(s, x^n), u(s', x^n)) \underset{(C_1)}{\leq} \frac{C'}{d(s, s')} d(u(s, x^n), u(s', x^n)) \end{aligned}$$

Im folgenden werden nun stets $(B_2) - (B_4)$, (C_1) , (C_2) und die gleichmäßige Ergodizität vorausgesetzt (welche nach (3.9) durch $(B_1) - (B_5)$ garantiert ist), ohne daß für jede einzelne Aussage wirklich alle Voraussetzungen benötigt werden.

Um die Beweisführungen übersichtlich gestalten zu können, wollen wir in (B_4) $n_0 = 1$ setzen.

(5.4) Lemma Setze

$A := \{x \in X : d(u(s, x), u(s', x)) \leq \lambda d(s, s') \text{ für alle } s, s' \in S\}$
(vgl. (B_4) und Bemerkung (5.3)) und

$$\lambda^n(x^n) := \sum_{i=1}^n \chi_A(x_i)$$

Dann gilt für alle $n \in \mathbb{N}$, $x^n \in X^n$, $s, s' \in S$

$$\begin{aligned} d(u(s, x^n), u(s', x^n)) & \leq \lambda^{\lambda^n(x^n)} d(s, s') \\ |p(u(s, x^n), x) - p(u(s', x^n), x)| & \leq C \lambda^{\lambda^n(x^n)} \end{aligned}$$

Beweis Folgt aus Definition (1.1) c, (B_3) , (B_4) , (C_2) .

(5.5) Definition Sei X ein polnischer Raum (polnisch = vollständig metrisierbar und separabel), $\mathcal{X}$ die zugehörigen Borelschen Teilmengen von X . Man definiert wie folgt ein Wahrscheinlichkeitsmaß $\tilde{P}$ auf $\mathcal{X}^{\mathbb{Z}}$:

Für jedes $n \in \mathbb{N}$ definiert man zuerst $\tilde{P}|_{\mathcal{X}^{[-n, \infty)}}$ durch:

Für $A \in \mathcal{X}^{[-n, \infty)}$ sei $\tilde{P}(A) := P(T^{-(n+1)}A)$

(wegen $T^{-(n+1)}A \in \mathcal{X}^{\mathbb{N}}$ ist die rechte Seite definiert.)

Dann ist $\tilde{P}$ die Kolmogorovsche Erweiterung der miteinander verträglichen Maße $\tilde{P}|_{\mathcal{X}^{[-n, \infty)}}$, $n \in \mathbb{N}$, auf $\mathcal{X}^{\mathbb{Z}}$. $\tilde{P}$ ist wieder stationär und $T: \mathcal{X}^{\mathbb{Z}} \rightarrow \mathcal{X}^{\mathbb{Z}}$ sowie T^{-1} bez. $\tilde{P}$ ergodisch (vgl. /6/, Th. 1.4 und p. 9). Wir schreiben im folgenden $P|_{\mathcal{X}^{\mathbb{Z}}}$ anstelle von $\tilde{P}|_{\mathcal{X}^{\mathbb{Z}}}$.

$x^{[-n, -1]} = (x_{-n}, \dots, x_{-1})$ soll im folgenden x^{-n} abgekürzt werden. x^{-m-n} steht dann zum Beispiel für $(x_{-m-n}, \dots, x_{-1})$.

(5.6) Lemma Für P-fast alle $x^{-\mathbb{N}} \in X^{-\mathbb{N}}$ gilt:

(a) Es existiert $u(x^{-\mathbb{N}}) := \lim_n u(s, x^{-n}) \in S$,
und es ist $u(x^{-\mathbb{N}})$ unabhängig von s .

(b) Es ist für jedes $n \in \mathbb{N}$:

$$u(x^{-n}) = u(s^*, x^{-n}) \quad \text{mit } s^* = u(x^{(-\infty, -(n+1))}) .$$

Beweis Zunächst kann man zeigen, daß für P-fast alle $x^{-\mathbb{N}}$

die Folge $(u(s, x^{-n}))_{n \in \mathbb{N}}$ eine Cauchyfolge ist. Wegen

$$u(s, x^{-m-n}) = u(s', x^{-n}) \quad \text{mit } s' = u(s, x^{[-m-n, -1-n]}) \text{ gilt}$$

$$d(u(s, x^{-m-n}), u(s, x^{-n})) \leq \lambda^{v^n(x^{-n})} \text{ gemäß (5.4) .}$$

Satz (3.18), modifiziert für den bzgl. $P|_{\mathcal{X}^{\mathbb{Z}}}$ ergodischen $\text{Op. } T^{-1}$

$$\text{(vgl. 5.5), liefert nun } \lim_n v^n(x^{-n}) = \infty \text{ für P-fast alle } x^{-\mathbb{N}} ,$$

also die Behauptung. Nach (C_1) ist S vollständig, also existiert ein $u(x^{-\mathbb{N}}) \in S$ als Limes der Cauchyfolge $u(s, x^{-n})$. Daß dieser

Limes unabhängig von s ist, folgt aus

$$d(u(s', x^{-n}), u(s'', x^{-n})) \leq \lambda^{v^n(x^{-n})} \text{ für alle } s', s'' \in S .$$

Behauptung (b) folgt nun so:

$$\begin{aligned} u(x^{-\mathbb{N}}) &= \lim_m u(s, x^{-m-n}) \\ &= \lim_m u(u(s, x^{[-m-n, -1-n]}), x^{-n}) \end{aligned}$$

$$\begin{aligned} &= \text{(Stetigkeit)} \quad u(\lim_m u(s, x^{[-m-n, -1-n]}), x^{-n}) = u(s^{\#}, x^{-n}) \\ \text{mit } s^{\#} &= \lim_m u(s, x^{[-m-n, -1-n]}) = u(x^{(-\infty, -(n+1)]}) \quad \text{w.z.b.w.} \end{aligned}$$

Zu $u(x^{-N}) \in S$ existiert das Wahrscheinlichkeitsmaß $P(u(x^{-N}), A)$
 $A \in \mathcal{X}$, mit der 1-Dichte $p(u(x^{-N}), x_0)$, die abgekürzt mit
 $P(x^{-N}, A) := P(u(x^{-N}), A)$
 $p(x^{-N}, x_0) := p(u(x^{-N}), x_0)$

bezeichnet werden soll.⁶ Aus Lemma (5.2) folgt dann sofort, daß
für jedes $n \in \mathbb{N}$ das Wahrscheinlichkeitsmaß

$$P(x^{-N}, A^{[0, n-1]}) := P(u(x^{-N}), A^{[0, n-1]}) \quad , \quad A^{[0, n-1]} \in \mathcal{X}^{[0, n-1]}$$

die 1ⁿ-Dichte

$$p(x^{-N}, x^{[0, n-1]}) := p(u(x^{-N}), x^{[0, n-1]}) = \prod_{k=0}^{n-1} p(u(x^{(-\infty, k-1]}), x_k))$$

hat.

(5.7) Lemma Für P-fast alle $x^{-N} \in X^{-N}$ gilt

- (a) $p(x^{-N}, x_0) = \lim_n p(u(s, x^{-n}), x_0)$ gleichmäßig in s ,
- (b) $p(x^{-N}, x_0)$ ist $\mathcal{X}^{-N} \times \mathcal{X}_0$ meßbar.

Beweis Mit Hilfe von (5.6) b liefert Lemma (5.4) die Ungleichung
 $|p(x^{-N}, x_0) - p(u(s, x^{-n}), x_0)| \leq C \lambda^{\nu^n(x^{-n})}$.

Es folgt wieder (wie im Beweis zu Lemma (5.6)) $\nu^n(x^{-n}) \rightarrow \infty$
für P-fast alle x^{-N} und damit Behauptung (a).

Nach (C₁) ist jedes $p(u(s, x^{-n}), x_0)$ $\mathcal{X}^{-n} \times \mathcal{X}_0$ meßbar, also auch
 $\mathcal{X}^{-N} \times \mathcal{X}_0$ meßbar. Dies gilt dann auch für den Limes. w.z.b.w.

(5.8) Satz Für jedes $A_0 \in \mathcal{X}$ ist P-fast überall

$$P(x^{-N}, A_0) = P(A_0 | \mathcal{X}^{-N}) (x^{-N})$$

Beweis Zu zeigen ist

$$\int_{A^{-N}} P(dx^{-N}) P(x^{-N}, A_0) = P(A^{-N} A_0)$$

für alle $A_0 \in \mathfrak{X}_0$ und für alle $A^{-N} \in \mathfrak{X}^{-N}$, oder -was bereits ausreicht- für alle $A^{-m} \in \mathfrak{X}^{-m}$, $m \in \mathbb{N}$.

Wegen (B_2) und (5.6) a gilt für P-fast alle x^{-N}

$$P(x^{-N}, A_0) = \lim_n P(u(s^*, x^{-n}), A_0) \quad \text{für jedes } s^* \in S$$

und damit wegen des Satzes von der majorisierten Konvergenz (Die Menge $A^{-m} \subset X^{-m}$ unter dem Integralzeichen soll stets als Teilmenge von X^{-N} verstanden werden in diesem Sinne:

$$\{ (\dots, x_{-2}, x_{-1}) \in X^{-N} : (x_{-m}, \dots, x_{-2}, x_{-1}) \in A^{-m} \}.$$

Wegen x^{-n} , x^{-m-n} usw. siehe die Erklärung vor Lemma (5.6))

$$\int_{A^{-m}} P(dx^{-N}) P(x^{-N}, A_0)$$

$$= \int_{A^{-m}} P(dx^{-N}) \lim_n P(u(s^*, x^{-n-m}), A_0)$$

$$= \lim_n \int_{A^{-m}} P(dx^{-n-m}) P(u(s^*, x^{-n-m}), A_0)$$

$$= \lim_n \lim_N \frac{1}{N} \sum_{k=1}^N \int_{A^{-m}} P(s, dx^{-k-n-m}) P(u(s^*, x^{-n-m}), A_0)$$

$$\text{In der letzten Gleichung haben wir } P(A^n) = \lim_N \frac{1}{N} \sum_{k=1}^N P^k(s, A^n)$$

für alle $s \in S$, $n \in \mathbb{N}$, $A^n \in \mathfrak{X}^n$ und die folgende Tatsache ausgenutzt:

$$(*) \left\{ \begin{array}{l} \text{Sind } \mu_n \text{ Wahrscheinlichkeitsmaße, } f \text{ eine meßbare Funktion,} \\ 0 \leq f \leq 1, \text{ dann:} \\ \lim_n \mu_n(A) = \mu(A) \quad \text{für alle } A \in \mathfrak{X} \Rightarrow \lim_n \int f d\mu_n = \int f d\mu \end{array} \right.$$

Andererseits ist wegen der Stationarität von P

$$P(A^{-m} A_0) = \lim P(X^{[-n-m, -1-m]} A^{-m} A_0)$$

$$= \lim_n \lim_N \frac{1}{N} \sum_{k=1}^N P(s, X^{[-k-n-m, -1-m]} A^{-m} A_0)$$

$$(1.8) \quad = \lim_n \lim_N \frac{1}{N} \sum_{k=1}^N \int_{A-m} P(s, dx^{-k-n-m}) P(u(s, x^{-k-n-m}), A_0)$$

Also haben wir wegen $u(s, x^{-k-n-m}) = u(s^{**}, x^{-n-m})$ mit

$$s^{**} = u(s, x^{[-k-n-m, -1-n-m]}) :$$

$$\left| \int_{A-m} P(dx^{-N}) P(x^{-N}, A_0) - P(A^{-m} A_0) \right|$$

$$\leq \lim_n \sup_N \lim_N \sup \frac{1}{N} \sum_{k=1}^N \int_{A-m} P(s, dx^{-k-n-m}) .$$

$$\cdot \left| P(u(s^{*}, x^{-n-m}), A_0) - P(u(s^{**}, x^{-n-m}), A_0) \right|$$

$$\leq B \lim_n \sup_N \lim_N \sup \frac{1}{N} \sum_{k=1}^N \int_{A-m} P(s, dx^{-k-n-m}) .$$

$$\cdot d(u(s^{*}, x^{-n-m}), u(s^{**}, x^{-n-m}))$$

$$\leq (5.4) \quad B \lim_n \sup_N \lim_N \sup \frac{1}{N} \sum_{k=1}^N \int_{A-m} P(s, dx^{-k-n-m}) \lambda^{n+m}(x^{-n-m})$$

$$= B \lim_n \sup \int_{A-m} P(dx^{-n-m}) \lambda^{n+m}(x^{-n-m})$$

$$= B \int_{X^{-N}} P(dx^{-N}) \lim_n \sup \lambda^{n+m}(x^{-n-m}) = 0 \quad (3.18)$$

(5.9) Bemerkung Schreibt man Satz (5.8) in der Form

$$P(A_0 | \mathcal{X}^{-N}) = \lim_n P(u(s, x^{-n}), A_0) \quad P\text{-fast überall,}$$

so kommt man in die Nähe eines Korollares zu den Doob'schen Martingalkonvergenzsätzen, das besagt,

$$P(A_0 | \mathcal{X}^{-N}) = \lim_n P(A_0 | \mathcal{X}^{-n}) \quad P\text{-fast überall}$$

(5.10) Korollar

(a) Für alle $A \in \mathcal{X}^{(-\infty, 0]}$ gilt

$$P(A) = \int_A P(x^{-N}, x_0) P(dx^{-N}) 1(dx_0)$$

(b) Für alle $A \in \mathcal{X}^{(-\infty, n]}$ gilt

$$P(A) = \int_A P(x^{(-\infty, n-1]}, x_n) P(dx^{(-\infty, n-1]}) 1(dx_n)$$

(c) Für alle $A \in \mathcal{X}^{(-\infty, n]}$ gilt

$$P(A) = \int P(x^{-N}, x^{[0, n]}) P(dx^{-N}) 1^{n+1}(dx^{[0, n]})$$

(d) Sei $f(x^{-N}, x_0)$ eine $P|_{\mathcal{X}^{(-\infty, 0]}}$ integrierbare Funktion.

Für den Erwartungswert $Ef = \int f dP$ bezüglich P gilt dann

$$Ef = \int_{\mathcal{X}^{(-\infty, 0]}} f(x^{-N}, x_0) P(x^{-N}, x_0) P(dx^{-N}) 1(dx_0)$$

Beweis

(a) Für alle $A^{-N} \in \mathcal{X}^{-N}, A_0 \in \mathcal{X}_0$ gilt wegen Satz (5.8)

$$P(A^{-N} A_0) = \int_{A_0} P(dx^{-N}) P(x^{-N}, x_0) 1(dx_0)$$

weil $P(x^{-N}, \cdot)|_{\mathcal{X}_0}$ die 1-Dichte $p(x^{-N}, \cdot)$ besitzt. Dann gilt

diese Gleichung auch für alle $A \in \mathcal{X}^{-N} \times \mathcal{X}_0 = \mathcal{X}^{(-\infty, 0]}$

(b) Folgt aus (a) und der Stationarität von P .

(c) Satz (5.8) läßt sich auch in der Form beweisen:

$$(*) P(x^{-N}, A^{[0, n]}) = P(A^{[0, n]} | \mathcal{X}^{-N}) \quad \text{für alle } A^{[0, n]} \text{ (P-fast über- all)}$$

Anstelle von (B_2) braucht man dann im Beweis folgende Tatsache, die leicht aus (B_2) , (B_3) herleitbar ist (vgl./13/, Lemma 2.1.63 case 1).

$$|P(s, A^m) - P(s', A^m)| \leq m B d(s, s') \quad \text{für alle } A^m \in \mathcal{X}^m, m \in \mathbb{N}.$$

Die Behauptung (c) folgt nun aus $(*)$ so, wie die Behauptung (a) aus (5.8) folgte, weil $p(x^{-N}, x^{[0, n]})$ die 1^{n+1} -Dichte von $P(x^{-N}, A^{[0, n]})$ ist (vgl. die Bemerkung unter (5.6)).

(d) Folgt sofort aus (a), wonach $p(x^{-N}, x_0)$ die $P|_{\mathcal{X}^{-N}} \otimes 1|_{\mathcal{X}_0}$ -Dichte von $P|_{(\mathcal{X}^{-N} \times \mathcal{X}_0)}$ ist.

Schließlich formulieren wir noch die Bedingung

$$(C_3) \quad E \log p(x^{-N}, x_0) < \infty \quad (E \text{ bezieht sich auf die Wahrscheinlichkeit } P|_{\mathcal{X}^{\mathbb{Z}}})$$

(5.11) Bemerkung Aufgrund von Korollar (5.10) d lautet

$$(C_3) : \int_{X(-\infty, 0]} p(x^{-N}, x_0) \log p(x^{-N}, x_0) P(dx^{-N}) 1(dx_0) < \infty$$

Wegen $p \log p \geq -\frac{1}{2}$ für $p \geq 0$ ist $E \log p(x^{-N}, x_0) \geq -\frac{1}{2}$.

(C_3) stellt eine Bedingung an die großen Werte von $p(x^{-N}, x_0)$ dar. (C_3) ist zum Beispiel erfüllt, wenn es ein $L < \infty$ gibt mit $\int p(s, x) \log p(s, x) 1(dx) \leq L$ oder $p(s, x) \leq L$ für alle $s \in S$.

Wir zeigen noch, wie die Voraussetzungen $(C_1) - (C_3)$ für eine lineare OM-Kette erfüllt werden können. (vgl. (2.6), (2.8))

(5.12) Satz Sei (X, S, P, u) eine lineare OM-Kette und $L < \infty$.

a) $S = \{\mu \in M_1 : \text{für die 1-Dichte } p(\mu, x) \text{ gilt}$

$\int p(\mu, x) \log p(\mu, x) 1(dx) \leq L\}$. Sei $\Lambda_x \in S$ für alle $x \in X$, $p(\Lambda_x, y)$ als Funktion von $(x, y) \in \mathcal{X}^2$ -meßbar und α_x als Fu. von x $\mathcal{X}$ -meßbar. Dann sind $(C_1), (C_3)$ erfüllt.

b) Setzt man in a) anstelle der Integralungleichung die Ungleichung $p(\mu, x) \leq L$, so sind sogar $(C_1), (C_2), (C_3)$ erfüllt.

Beweis Zu (C_1) : Wir zeigen zuerst, daß S im Fall a) vollständig ist. Der Fall b) verläuft dann analog.

Sei also μ_n eine Cauchyfolge aus S und $\mu = \lim_n \mu_n$, $\mu \in M_1$ (M_1 ist vollständig). Wegen (2.8) gilt

$$0 = \lim_n \sup_{A \in \mathcal{X}} |\mu(A) - \mu_n(A)| = \lim_n \frac{1}{2} \int |p(\mu_n, x) - p(\mu, x)| 1(dx)$$

Nach /1/, Satz 15.5, existiert nun eine Teilfolge n_i mit

$$p(\mu, x) = \lim p(\mu_{n_i}, x) \quad \text{für 1-fast alle } x \in X$$

und damit auch wegen der Stetigkeit der Funktion $p \log p$ für $p \geq 0$

$$p(\mu, x) \log p(\mu, x) = \lim p(\mu_{n_i}, x) \log p(\mu_{n_i}, x) \quad \text{für 1-fast alle } x.$$

Mit Hilfe des Lemmas von Fatou (/1/, Lemma 15.1) schließt man nun aus

$\int p(\mu_{n_1}, x) \log p(\mu_{n_1}, x) \, l(dx) \leq L$ für alle n_1 auf

$\int p(\mu, x) \log p(\mu, x) \, l(dx) \leq L$,

womit S als vollständig nachgewiesen ist.

Daß S auch wirklich als Zustandsraum geeignet ist, wird aber erst durch

$\mu \in S \Rightarrow u(\mu, x) \in S$ für alle $x \in X$

gezeigt. Sei also $\mu \in S$, $x \in X$. Es gilt

$$p(u(\mu, x), y) = \alpha_x p(\mu, y) + (1 - \alpha_x) p(\Lambda_x, y)$$

und deshalb mit Hilfe der Konvexität von

$p \log p$ für $p \geq 0$:

$$\int p(u(\mu, x), y) \log p(u(\mu, x), y) \, l(dy) \leq$$

$$\alpha_x \int p(\mu, y) \log p(\mu, y) \, l(dy) + (1 - \alpha_x) \int p(\Lambda_x, y) \log p(\Lambda_x, y) \, l(dy) \leq L,$$

also $u(\mu, x) \in S$.

Aus Lemma (2.11) folgt ferner, daß

$$(*) \quad p(u(\mu, x^n), x) = \alpha_{x_n} \dots \alpha_{x_1} p(\mu, x) + \alpha_{x_n} \dots \alpha_{x_2} (1 - \alpha_{x_1}) p(\Lambda_{x_1}, x) \\ + \dots + (1 - \alpha_{x_n}) p(\Lambda_{x_n}, x),$$

also auch die $\mathcal{K}^n \times \mathcal{K}$ -Meßbarkeit von $p(u(\mu, x^n), x)$.

Zu (C_2) : Im Fall b) liefert (*) einerseits

$$|p(u(\mu, x^n), x) - p(u(\mu', x^n), x)| \leq \alpha_{x_1} \dots \alpha_{x_n} 2L.$$

Andererseits folgt aus Lemma (2.11)

$$d(u(\mu, x^n), u(\mu', x^n)) = \alpha_{x_1} \dots \alpha_{x_n} d(\mu, \mu'),$$

womit (C_2) gültig ist.

Zu (C_3) : In a) und b) ist gemäß Bemerkung (5.11) Bedingung

(C_3) erfüllt.

6. Eine Version des Shannon-McMillan Theorems

Das Ziel dieses Paragraphen ist es, für eine gleichmäßig ergodische Kette mvB die $(P_S$ -fast überall)-Konvergenz von $\frac{1}{n} \log p(s, x^n)$ bei $n \rightarrow \infty$ zu beweisen, wobei $p(s, x^n)$ die Dichten der (nicht stationären) Wahrscheinlichkeiten $P(s, A^n)$, $A^n \in \mathcal{X}^n$, sind. Aus dem klassischen Beweis des Shannon-McMillan Theorems (siehe/6/, S. 129, 13.1 für endliches X und /15/ für beliebiges X) in welchem die $(P$ -fast überall)-Konvergenz von $\frac{1}{n} \log p(x^n)$ betrachtet wird, wobei die $p(x^n)$ Dichten der (stationären) Wahrscheinlichkeit $P|_{\mathcal{X}^{\mathbb{N}}}$ sind, übernehmen wir die Idee, $P|_{\mathcal{X}^{\mathbb{N}}}$ auf $P|_{\mathcal{X}^{\mathbb{Z}}}$ (vgl. Definition (5.5)) und $p(s, x_0)$ auf eine Funktion $p(x^{-N}, x_0)$ (vgl. Lemma (5.7)) zu erweitern. Es wird wieder vorausgesetzt, daß die Kette mvB ergodisch ist und die Voraussetzungen $(B_2) - (B_4)$, $(C_1) - (C_3)$ erfüllt.

(6.1) Lemma Es gibt ein $A_1 \in \mathcal{X}^{\mathbb{N}}$ mit $P(A_1) = 1$ und Zahlen $C < \infty$, $\beta < 1$, so daß für alle $x^{\mathbb{N}} \in A_1$ ein $N = N(x^{\mathbb{N}})$ existiert mit

$$|p(u(s', x^{n-1}), x_n) - p(u(s'', x^{n-1}), x_n)| \leq C \beta^n \quad \text{für alle } n \geq N, s', s'' \in S.$$

Beweis Lemma (5.4) liefert (Setze in (5.4) $C \cdot \lambda$ anstelle von C) $|p(u(s', x^{n-1}), x_n) - p(u(s'', x^{n-1}), x_n)| \leq C \lambda^{v^n(x^n)}$.

Nach Satz (3.18) existiert ein $A_1 \in \mathcal{X}^{\mathbb{N}}$ mit $P(A_1) = 1$, so daß es für alle $x^{\mathbb{N}} \in A_1$ ein $N = N(x^{\mathbb{N}})$ gibt mit $v^n(x^n) \geq \alpha n$ für alle $n \geq N$ ($\alpha > 0$).

Mit $\beta := \lambda^\alpha < 1$ ist dann die Behauptung bewiesen.

(6.2) Lemma Sei $\varepsilon_n \geq 0$ eine Folge mit $\sum_{n=1}^{\infty} \varepsilon_n < \infty$. Dann existiert ein $A_2 \in \mathcal{X}^{\mathbb{N}}$ mit $P(A_2) = 1$, für das gilt:

Für jedes $x^{\mathbb{N}} \in A_2$ gibt es ein $N = N(x^{\mathbb{N}})$ und ein $x^{-N} \in X^{-N}$,
so daß $p(x^{-N}, x_0)$ existiert und für alle $n \geq N$ gilt:
 $p(x^{(-\infty, n-1]}, x_n) \geq \varepsilon_n$.

Beweis Für P-fast alle $x^{\mathbb{Z}} = (x^{-N}, x_0, x^{\mathbb{N}}) \in X^{\mathbb{Z}}$ existiert
 $u(x^{-N}) \in S$ nach Lemma (5.6) und damit auch $p(x^{-N}, x_0)$. Setze für
diese $x^{\mathbb{Z}} \in X^{\mathbb{Z}}$:

$p(x^{\mathbb{Z}}) := p(x^{-N}, x_0)$. Dann ist für jedes $n \in \mathbb{N}$
 $p(T^n x^{\mathbb{Z}}) = p(x^{(-\infty, n-1]}, x_n)$ ebenfalls existent. Aufgrund von
(5.10) b gilt

$$P\{p(T^n x^{\mathbb{Z}}) < \varepsilon_n\} = \int_{\{p(T^n x^{\mathbb{Z}}) < \varepsilon_n\}} p(T^n x^{\mathbb{Z}}) p(dx^{(-\infty, n-1]})_1(dx_n) \leq \varepsilon_n$$

und deshalb

$$\sum_{n=1}^{\infty} P\{p(T^n x^{\mathbb{Z}}) < \varepsilon_n\} \leq \sum_{n=1}^{\infty} \varepsilon_n < \infty.$$

Aus dem Borel-Cantelli Lemma folgt die Existenz eines $\tilde{A}_2 \in \mathcal{X}^{\mathbb{Z}}$
mit $P(\tilde{A}_2) = 1$, so daß für alle $x^{\mathbb{Z}} \in \tilde{A}_2$ das Ereignis
 $p(T^n x^{\mathbb{Z}}) < \varepsilon_n$ nur für endlich viele n eintritt, oder anders
formuliert, daß es ein $N = N(x^{\mathbb{Z}})$ gibt mit $p(T^n x^{\mathbb{Z}}) \geq \varepsilon_n$
für alle $n \geq N$.

Da X als polnisch vorausgesetzt wurde (vgl. (5.5)), gibt es ein
 $A_2 \subset \text{proj}_{X^{\mathbb{N}}} \tilde{A}_2$ mit $A_2 \in \mathcal{X}^{\mathbb{N}}$ und $P(A_2) = 1$, womit das Lemma bewiesen
ist.

(6.3) Lemma Es existiert ein $A_3 \in \mathcal{X}^{\mathbb{N}}$ mit $P(A_3) = 1$, so daß es
für alle $x^{\mathbb{N}} \in A_3$ ein $N = N(x^{\mathbb{N}})$ gibt mit

- (a) $p(u(s, x^{n-1}), x_n) > 0$ für alle $n \geq N$ und $s \in S$
- (b) $|\log p(u(s', x^{n-1}), x_n) - \log p(u(s'', x^{n-1}), x_n)| \rightarrow 0$
bei $n \rightarrow \infty$ für alle $s', s'' \in S$.

Beweis Gemäß Lemma (6.1) gibt es für alle $x^N \in A_1$ ein N_1 mit $p(u(s, x^{n-1}), x_n) \geq p(u(s^*, x^{n-1}), x_n) - C\beta^n$

für alle $s, s^* \in S$, $n \geq N_1$. Gemäß Lemma (6.2) gibt es zur Folge $\tilde{\beta}^n$ ($\beta < \tilde{\beta} < 1$) ein $A_2 \in \mathcal{X}^N$ mit $P(A_2) = 1$, so daß es für alle $x^N \in A_2$ ein N_2 und ein $s^* = u(x^{(-\infty, 0]}) \in S$ gibt mit $p(u(s^*, x^{n-1}), x_n) \geq C \tilde{\beta}^n$ für alle $n \geq N_2$.

Setze nun $A_3 := A_1 \cap A_2$. Es ist $P(A_3) = 1$. Für alle $x^N \in A_3$ und $n \geq \max(N_1, N_2)$ ist dann für alle $s \in S$

$$p(u(s, x^{n-1}), x_n) \geq C (\tilde{\beta}^n - \beta^n),$$

womit bereits (a) bewiesen ist.

Zum Beweis von Teil (b) benutzen wir die Ungleichung

$$|\log a - \log b| \leq \frac{|a - b|}{a} + \frac{|a - b|}{b} \quad (a, b > 0)$$

die aus der Ungleichung $\log \frac{a}{b} \leq \frac{a}{b} - 1$ folgt.

Für $x^N \in A_3$ und $n \geq \max(N_1, N_2)$ gilt nun für alle $s', s'' \in S$:

$$\begin{aligned} & |\log p(u(s', x^{n-1}), x_n) - \log p(u(s'', x^{n-1}), x_n)| \\ & \leq \frac{|p(u(s', x^{n-1}), x_n) - p(u(s'', x^{n-1}), x_n)|}{p(u(s', x^{n-1}), x_n)} \\ & + \frac{|p(u(s', x^{n-1}), x_n) - p(u(s'', x^{n-1}), x_n)|}{p(u(s'', x^{n-1}), x_n)} \\ & \leq \frac{2\beta^n}{\tilde{\beta}^n - \beta^n} \longrightarrow 0 \text{ bei } n \rightarrow \infty. \quad \text{w.z.b.w.} \end{aligned}$$

(6.4) Lemma Es existiert ein $A_4 \in \mathcal{X}^N$ mit $P(A_4) = 1$, so daß es für alle $x^N \in A_4$ ein $x^{-N} \in X^{-N}$ gibt mit:

$p(x^{-N}, x_0)$ existiert,

$p(x^{(-\infty, n-1]}, x_n) > 0$ für alle $n \in \mathbb{N}$ und

$$\lim_n \frac{1}{n} \sum_{k=1}^n \log p(x^{(-\infty, k-1]}, x_k) = E \log p(x^{-N}, x_0)$$

(E bezieht sich wie immer auf $P|_{\mathcal{X}^{\mathbb{Z}}}$).

Beweis Wie im Beweis zu (6.2) setzen wir

$p(x^{\mathbb{Z}}) := p(x^{-N}, x_0)$, was für P-fast alle $x^{\mathbb{Z}}$ existiert. Nun gilt $P\{x^{\mathbb{Z}} : \text{es ex. ein } n \in \mathbb{N} \text{ mit } p(T^n x^{\mathbb{Z}}) = 0\} \leq \sum_n P\{p(T^n x^{\mathbb{Z}}) = 0\} = 0$ wegen (5.10) b, also für P-fast alle $x^{\mathbb{Z}} : p(T^n x^{\mathbb{Z}}) > 0$ für alle $n \in \mathbb{N}$. Gemäß Voraussetzung (C_3) ist $\log p(x^{\mathbb{Z}})$ P-integrierbar und damit das in (3.15) zitierte individuelle Ergodentheorem gültig. Es gibt also ein $\tilde{A}_4 \in \mathcal{X}^{\mathbb{Z}}$ mit $P(\tilde{A}_4) = 1$, so daß für alle $x^{\mathbb{Z}} \in \tilde{A}_4$ $p(T^k x^{\mathbb{Z}}) > 0$ ($k \in \mathbb{N}$) und

$$\lim_n \frac{1}{n} \sum_{k=1}^n \log p(T^k x^{\mathbb{Z}}) = E \log p(x^{\mathbb{Z}}) \quad \text{gilt.}$$

Da X als polnisch vorausgesetzt ist (vgl. (5.5)), gibt es ein

$A_4 \subset \text{proj}_{X^{\mathbb{N}}} \tilde{A}_4$ mit $A_4 \in \mathcal{X}^{\mathbb{N}}$ und $P(A_4) = 1$, womit das Lemma bewiesen ist.

Nun kommt die angekündigte Version des Shannon-McMillan Theorems. Wir setzen dabei

$$h := E \log p(x^{-N}, x_0)$$

Es ist $|h| < \infty$ gemäß Voraussetzung (C_3) . Weiter setzen wir zur Abkürzung

$$p(sx^{k-1}, x_k) := p(u(s, x^{k-1}), x_k) .$$

(6.5) Satz Zu jedem $s \in S$ existiert ein $A \in \mathcal{X}^{\mathbb{N}}$ mit $P_s(A) = 1$ so daß für alle $x^{\mathbb{N}} \in A$ gilt:

$$p(s, x^n) > 0 \quad \text{für alle } n \in \mathbb{N} \text{ und}$$

$$\lim_n \frac{1}{n} \log p(s, x^n) = h$$

Beweis Gemäß Lemma (5.2) gilt

$$\log p(s, x^n) = \sum_{k=1}^n \log p(sx^{k-1}, x_k) .$$

Setze $A_5 := A_3 \cap A_4 \in \mathcal{X}^{\mathbb{N}}$ (vgl. (6.3), (6.4)). Es ist $P(A_5) = 1$. Sei nun $x^{\mathbb{N}} \in A_5$.

Gemäß Lemma (6.3) a existiert ein $N \in \mathbb{N}$ mit

$$(*) \quad p(sx^{n-1}, x_n) > 0 \quad \text{für alle } n \geq N, s \in S.$$

Gemäß Lemma (6.4) existiert ein $s^* = u(x^{-N}) \in S$ mit

$$p(s^* x^{n-1}, x_n) > 0 \quad \text{für alle } n \in \mathbb{N} \text{ und}$$

$$\lim_n \frac{1}{n} \sum_{k=1}^n \log p(s^* x^{k-1}, x_k) = h.$$

Daraus folgt für alle $n \in \mathbb{N}$

$$\left| \frac{1}{n} \sum_{k=N}^n \log p(sx^{k-1}, x_k) - \frac{1}{n} \sum_{k=N}^n \log p(s^* x^{k-1}, x_k) \right|$$

$$\leq \frac{1}{n} \sum_{k=N}^n \left| \log p(sx^{k-1}, x_k) - \log p(s^* x^{k-1}, x_k) \right| \rightarrow 0 \quad \text{für } n \rightarrow \infty$$

nach Lemma (6.3) b und deshalb

$$(**) \quad \lim_n \frac{1}{n} \sum_{k=N}^n \log p(sx^{k-1}, x_k) = h \quad \text{für alle } s \in S.$$

Nun wollen wir von der Konvergenz (P-fast überall) auf die Konvergenz (P_S -fast überall) schließen. Dazu setzen wir

$$A^* := \lim_n \sup T^{-n} A_5 = \bigcap_{m \geq 0} \bigcup_{n \geq m} T^{-n} A_5$$

Es ist $A^* \in \mathfrak{A}^{\mathbb{N}}$, und es gilt

$$(i) \quad P(A^*) = 1; \text{ denn es ist}$$

$$P(A^*) = \lim_m P\left(\bigcup_{n \geq m} T^{-n} A_5\right) \geq \lim_m P(T^{-m} A_5) = P(A_5) = 1$$

(ii) A^* ist ein terminales Ereignis; denn für jedes $M \in \mathbb{N}$ ist

$$A^* = \bigcap_{m \geq M} \bigcup_{n \geq m} T^{-n} A_5 \in \mathfrak{A}^{[M, \infty)}$$

Aufgrund von (i) und (ii) liefert Satz (3.17) nun $P_S(A^*) = 1$ für alle $s \in S$.

(iii) Für jedes $x^{\mathbb{N}} \in A^*$ gibt es ein $N \in \mathbb{N}$, so daß (*) und (**) gelten.

Denn aus $x^{\mathbb{N}} \in A^*$ folgt $x^{\mathbb{N}} \in T^{-m} A_5$ für ein $m \in \mathbb{N}$. Setzt man $x^{\mathbb{N}} = (x^m, x^{[m+1, \infty)})$, so heißt das $x^{[m+1, \infty)} \in A_5$.

Nach Definition von A_S existiert demnach ein $N \in \mathbb{N}$ mit

$$p(\tilde{s}x^{[m+1, m+n-1]}, x_{m+n}) > 0 \quad \text{für alle } n \geq N \text{ und für alle } \tilde{s} \in S \text{ und}$$

$$\lim_n \frac{1}{n} \sum_{k=N}^n \log p(\tilde{s}x^{[m+1, m+k-1]}, x_{m+k}) = h \quad \text{für alle } \tilde{s} \in S.$$

Setze nun $\tilde{s} := u(s, x^m)$. Dann folgt wegen

$$u(\tilde{s}, x^{[m+1, m+n]}) = u(s, x^{m+n})$$

$$p(sx^{n-1}, x_n) > 0 \quad \text{für alle } n \geq N + m \text{ und für alle } s \in S \text{ und}$$

$$\lim_n \frac{1}{n} \sum_{k=N+m}^n \log p(sx^{k-1}, x_k) = h \quad \text{für alle } s \in S,$$

womit Behauptung (iii) bewiesen ist.

Sei nun ein $s \in S$ fixiert. Es existiert ein $A_s \in \mathcal{D}^{\mathbb{N}}$ mit

$$P_s(A_s) = 1, \text{ so daß für alle } x^{\mathbb{N}} \in A_s \text{ gilt:}$$

$$p(sx^{n-1}, x_n) > 0 \quad \text{für alle } n \in \mathbb{N}.$$

$$\text{Denn es ist } P_s \{ x^{\mathbb{N}} : \text{es gibt ein } n \text{ mit } p(sx^{n-1}, x_n) = 0 \}$$

$$\leq \sum_{n \in \mathbb{N}} P_s \{ x^{\mathbb{N}} : p(sx^{n-1}, x_n) = 0 \} = 0 \text{ aufgrund von Lemma (5.2).}$$

$$\text{Setze nun } A := A^* \cap A_s.$$

$$\text{Es ist } P_s(A) = 1 \quad \text{und für alle } x^{\mathbb{N}} \in A \text{ gilt:}$$

$$p(sx^{n-1}, x_n) > 0 \quad \text{für alle } n \in \mathbb{N} \text{ und}$$

$$\lim_n \frac{1}{n} \sum_{k=1}^n \log p(sx^{k-1}, x_k) = h,$$

womit der Satz bewiesen ist.

(6.6) Bemerkungen

1. Handelt es sich hier bei der Kette mvB um den Spezialfall eines Markovprozesses, so reduzieren sich die Ausführungen der Paragraphen 5 und 6 auf die Aussage des Lemma (6.4), welches dann nichts anderes als die Anwendung des individuellen Ergodentheorems auf die Funktion $\log p(x_{-1}, x_0)$ darstellt (vgl. /23/, (3.3) und /2/, Theorem 3.2).

2. Sei X endlich. Dann ist für $x_0 \in X$

$$p(x^{-N}, x_0) = P(x^{-N}, x_0) .$$

Gemäß /6/, S. 132, ist

$$E \log P(x^{-N}, x_0) = -H ,$$

wobei H die Entropie des Verschiebungsoperators T bezüglich der stationären Wahrscheinlichkeit $P|_{\mathcal{X}^{\mathbb{N}}}$ ist. (Die Definition der Entropie H für beliebiges X findet man in /6/, S. 61, (5.14)) Die Größe $h = E \log p(x^{-N}, x_0)$ ist möglicherweise ein Ansatzpunkt, um die für manche Fragestellungen sehr nützliche Informationstheorie (vgl. /6/, Chap.5) auf Ketten mvB anzuwenden.

7. Abhängigkeit der Kette mvB von einem Parameter ϑ und
Konsistenz der Maximum-Likelihood-Schätzung

Eine endliche lineare OM-Kette wird gemäß (4.1) durch die Festlegung des endlich-dimensionalen Parameters $\vartheta = (\alpha, \wedge)$ spezifiziert. Läßt sich nun ein in der Natur vorkommender stochastischer Prozeß als eine endliche lineare OM-Kette darstellen, so ist das "richtige" ϑ , dasjenige ϑ also, das der Realisierung der Kette zugrunde liegt, unbekannt. Es muß aus einer Stichprobe $x^n = (x_1, x_2, \dots, x_n) \in X^n$ geschätzt werden. Wir setzen nun auch im allgemeinen Fall voraus, daß die Kette mvB (X, S, P, u) von einem endlichdimensionalen reelwertigen Parameter ϑ abhängig ist. Mit Hilfe einiger Bedingungen über die Art der ϑ -Abhängigkeit kann man die Ergebnisse der beiden letzten Paragraphen auf diese neue Situation übertragen.

(7.1) Erläuterungen In einer Kette mvB mögen die Abbildungen $u(s, x, \vartheta)$, $s \in S$, $x \in X$ und die Wahrscheinlichkeiten $P(s, A, \vartheta)$, $s \in S$, $A \in \mathcal{K}$ von einem $\vartheta \in \Theta$, $\Theta \subset \mathbb{R}^m$, $m \in \mathbb{N}$ abhängen. Für jedes $\vartheta \in \Theta$ sollen dabei die Voraussetzungen in Definition (1.1) erfüllt sein.

Zu $\vartheta \in \Theta$ und $s \in S$ gehört die Wahrscheinlichkeit $P_s(\vartheta) \in \mathcal{K}^N$. Liegt einer Realisierung der Kette mvB für jedes $s \in S$ die Wahrscheinlichkeit $P_s(\vartheta_0) \in \mathcal{K}^N$ zugrunde, so soll (der nicht notwendig eindeutige Wert) $\vartheta_0 \in \Theta$ wahrer Parameter heißen. Wie bisher schreiben wir P_s anstelle von $P_s(\vartheta_0)$.

(7.2) Voraussetzungen Für das Folgende mögen die Voraussetzungen $(B_2) - (B_4)$, $(C_1) - (C_2)$, falls man in ihnen $P(s, A)$, $u(s, x)$, $p(s, x)$ und die daraus abgeleiteten Größen durch $P(s, A, \vartheta)$, $u(s, x, \vartheta)$, $p(s, x, \vartheta)$ und die entsprechend abgeleiteten Größen auswechselt, für jedes $\vartheta \in \Theta$ gültig sein. (Die in diesen Voraussetzungen auftretenden Konstanten $B, \gamma, \lambda, n_0, C$ sollen nicht von ϑ abhängen.) Ferner gelte die gleichmäßige Ergodizität für jedes $\vartheta \in \Theta$. (Es existiert dann für jedes $\vartheta \in \Theta$ eine stationäre Wahrscheinlichkeit $P(\vartheta) \in \mathcal{M}^{\mathbb{Z}}$. Anstelle von $P(\vartheta_0) \in \mathcal{M}^{\mathbb{Z}}$ schreiben wir wieder $P \in \mathcal{M}^{\mathbb{Z}}$.)

(7.3) Abkürzungen Wir setzen für jedes $\vartheta \in \Theta$, $n \in \mathbb{N}$, $x^n \in X^n$, $s \in S$

$$u(s, x^n, \vartheta) := u(u(\dots u(u(s, x_1, \vartheta), x_2, \vartheta) \dots x_{n-1}, \vartheta), x_n, \vartheta)$$

(vgl. Definition (1.1) c) und

$$p(s, x^{n-1}, x_n, \vartheta) := p(u(s, x^{n-1}, \vartheta), x_n, \vartheta)$$

(Gemäß (5.2) ist dann auch ein $p(s, x^n, \vartheta)$ definiert.)

(7.4) Bemerkungen Für eine OM-Kette kann $P(\mu, A) = \mu(A)$ nicht mehr von einem ϑ abhängen. Die ϑ -Abhängigkeit geht also bei einer OM-Kette nur über $u(\mu, x, \vartheta)$ ein.

Wie folgt, lassen sich nun die Aussagen des Paragraphen 5 übertragen. Dabei müssen $P(s, A)$, $u(s, x)$, $p(s, x)$ und die daraus abgeleiteten Größen -wo immer sie auftauchen- mit einem ϑ versehen werden. "P-fast alle" bedeutet -wie schon in (7.2) angedeutet- " $P(\vartheta_0)$ -fast alle".

(7.5) Lemma (5.1) - (5.10) , (5.12) gilt für jedes $\vartheta \in \Theta$.
 Insbesondere existieren für P-fast alle $x^{-N} \in X^{-N}$ die Größen
 $u(x^{-N}, \vartheta)$, $P(x^{-N}, A, \vartheta)$, und es gilt gleichmäßig in $s \in S$, $\vartheta \in \Theta$
 $p(x^{-N}, x_0, \vartheta) = \lim_n p(sx^{-n}, x_0, \vartheta)$.

Beweis folgt aus den Voraussetzungen (7.2) mit den gleichen
 Beweisen wie in Paragraph 5.

Nun brauchen wir noch drei spezielle Bedingungen über die Art
 der ϑ -Abhängigkeit.

(D₁) Für jedes $n \in \mathbb{N}$, $s \in S$, $x^n \in X^n$ ist

$p(sx^{n-1}, x_n, \vartheta)$ stetig in $\vartheta \in \Theta$.

(D₂) Für jedes $n \in \mathbb{N}$, $s \in S$ ist die Menge

$\{ x^n : p(sx^{n-1}, x_n, \vartheta) = 0 \} \in \mathcal{B}^n$ nicht von ϑ abhängig.

(7.6) Bemerkungen Die Bedingung (C₃) lautet, wenn $p(x^{-N}, x_0)$
 mit einem ϑ versehen wird

$$E \log p(x^{-N}, x_0, \vartheta) = \int p(x^{-N}, x_0, \vartheta_0) \log p(x^{-N}, x_0, \vartheta) P(dx^{-N}) 1(dx) < \infty$$

Man beachte aber, daß $E \log p(x^{-N}, x_0, \vartheta)$ den Wert $-\infty$ annehmen

kann. Wir brauchen deshalb noch zusätzlich eine Bedingung für
 die kleinen Werte von $p(x^{-N}, x_0, \vartheta)$. Dabei wird

$\log^- p := -\log p$ für $0 \leq p \leq 1$ und $:= 0$ für $p > 1$ gesetzt.

(D₃)* Für jedes $\vartheta \in \Theta$ gilt

$E \log p(x^{-N}, x_0, \vartheta) < \infty$ und

$E (\log^- p(x^{-N}, x_0, \vartheta))^\alpha < \infty$ mit einem $\alpha > 1$.

Nun übertragen wir den Paragraphen 6 auf den Fall, daß jedes
 $p(s, x)$, $u(s, x)$ mit einem ϑ versehen wird. P , P_s , E beziehen
 sich wieder auf den Wert $\vartheta_0 \in \Theta$.

(7.7) Lemma Die Lemmata (6.1) - (6.5) gelten für jedes $\vartheta \in \Theta$.
 (Die dort auftretenden $A_2, A_3, A_4, A \in \mathcal{X}^{\mathbb{N}}$ hängen noch von ϑ ab.) In Lemma (6.2) sind allerdings nur solche Folgen $\varepsilon_n \geq 0$ zugelassen, für welche $\sum_{n=1}^{\infty} (\log^- \varepsilon_n)^{-\alpha} < \infty$ sind (α wie in $(D_3)^*$, $(\beta^n)_n, \beta < 1$, ist zum Beispiel eine solche Folge).

Beweis (6.1) folgt wieder aus (5.5). In (6.2) bedarf es dieser Modifikation: Aus der Abschätzung

$$\int (\log^- p)^{\alpha} dP \geq \int_{\{p < \varepsilon\}} (\log^- p)^{\alpha} dP \geq (\log^- \varepsilon)^{\alpha} P(p < \varepsilon) \text{ folgt}$$

$$P\{p(T^n x^{\mathbb{Z}}, \vartheta) < \varepsilon_n\} \leq (\log^- \varepsilon_n)^{-\alpha} C, \text{ wobei}$$

$$C := E(\log^- p(x^{-N}, x_0, \vartheta))^{\alpha} < \infty \text{ gesetzt wurde.}$$

Wegen $\sum_n (\log^- \varepsilon_n)^{-\alpha} C < \infty$ ist also wieder das Borel-Catelli Lemma anwendbar.

(6.3) folgt wieder aus (6.1), (6.2). Der Beweis von (6.4)

ist direkt übertragbar, wenn man das nachfolgende Lemma

(7.8) beachtet. (6.5) folgt wieder aus (6.3), (6.4), wenn man (D_2) beachtet.

(7.8) Lemma Die Menge

$$\{x^{(-\infty, 0]} \in X^{(-\infty, 0]} : p(x^{-N}, x_0, \vartheta) = 0\}$$

ist nicht von ϑ abhängig.

(Man beachte, daß nach (7.5) für P-fast alle x^{-N} $p(x^{-N}, x_0, \vartheta)$ existiert, und zwar jeweils für alle ϑ gleichzeitig).

Beweis Für ein $x^{(-\infty, 0]}$ und ein $\tilde{\vartheta} \in \Theta$ sei

$p(x^{-N}, x_0, \tilde{\vartheta}) = 0$. Sei $\varepsilon > 0$ beliebig. Wir zeigen, daß dann für ein beliebiges $\vartheta \in \Theta$ gilt:

$$p(x^{-N}, x_0, \vartheta) \leq \varepsilon.$$

Zufolge (7.5) kann man ein $n \in \mathbb{N}$ wählen, so daß für alle $s \in S$

$|p(x^{-N}, x_0, \vartheta) - p(sx^{-N}, x_0, \vartheta)| \leq \varepsilon$ gilt.

Nun ist $p(x^{-N}, x_0, \tilde{\vartheta}) = p(\tilde{s}x^{-N}, x_0, \tilde{\vartheta}) = 0$ mit

$\tilde{s} := u(x^{(-\infty, -(n+1)]}, \tilde{\vartheta})$. Daraus folgt wegen (D_2) auch

$p(\tilde{s}x^{-N}, x_0, \vartheta) = 0$ und daraus (wegen der obigen Ungleichung,

setzt man in ihr speziell $s := \tilde{s}$):

$$p(x^{-N}, x_0, \vartheta) \leq \varepsilon$$

w.z.b.w.

Nach Lemma (7.7) konvergiert $\frac{1}{n} \log p(s, x^n, \vartheta)$ gegen

$E \log p(x^{-N}, x_0, \vartheta)$. Mit diesem Ausdruck wollen wir uns nun näher beschäftigen.

(7.9) Definition Wir setzen $h(\vartheta) := E \log p(x^{-N}, x_0, \vartheta)$.

(Wegen Bedingung $(D_3)^*$ ist $|h(\vartheta)| < \infty$ für jedes $\vartheta \in \Theta$.)

Der nächste Satz, den man auch mit entsprechend einfacheren Beweisen bei /23/, /27/ und /2/ findet, gibt uns die Möglichkeit, $P(\vartheta) | \mathcal{K}^{\mathbb{Z}}$ von $P(\vartheta_0) | \mathcal{K}^{\mathbb{Z}}$ mit Hilfe der Funktion $h(\vartheta)$ unterscheiden zu können.

(7.10) Satz Es gilt $h(\vartheta) \leq h(\vartheta_0)$. Das Gleichheitszeichen steht genau dann, wenn $P(\vartheta) | \mathcal{K}^{\mathbb{Z}} = P(\vartheta_0) | \mathcal{K}^{\mathbb{Z}}$.

Beweis Es ist

$$h(\vartheta) - h(\vartheta_0) = E \log \frac{p(x^{-N}, x_0, \vartheta)}{p(x^{-N}, x_0, \vartheta_0)}$$

$$= \int \log \frac{p(x^{-N}, x_0, \vartheta)}{p(x^{-N}, x_0, \vartheta_0)} p(x^{-N}, x_0, \vartheta_0) P(dx^{-N}) \quad (5.10) \quad d) \quad \int_{X_0} \log \frac{p(x^{-N}, x_0, \vartheta)}{p(x^{-N}, x_0, \vartheta_0)} p(x^{-N}, x_0, \vartheta_0) P(dx^{-N}) \quad 1(dx_0)$$

$$\leq \log \int_{X^{-N}} \left\{ \int_{X_0} p(x^{-N}, x_0, \vartheta) \quad 1(dx_0) \right\} P(dx^{-N}) = \log 1 = 0. \quad (\text{Jensen})$$

Die Jensensche Ungleichung liefert zudem, daß das Gleichheitszeichen genau dann gilt, wenn es ein $\tilde{\Delta}^0 \in \mathcal{K}^{(-\infty, 0]}$ gibt mit

$P(\tilde{A}^0) = 1$, so daß für alle $(x^{-N}, x_0) \in \tilde{A}^0$ gilt:

$$p(x^{-N}, x_0, \mathcal{V}) = p(x^{-N}, x_0, \mathcal{V}_0)$$

Bilde für jedes $n \in \mathbb{N}$

$$\tilde{A}^n := \bigcap_{k=0}^n T^{-k} \tilde{A}^0 \in \mathfrak{D} \in (-\infty, n], \quad P(\tilde{A}^n) = 1.$$

Für alle $x^{(-\infty, n]} \in \tilde{A}^n$ gilt für $k = 0, 1, \dots, n$.

$$p(x^{(-\infty, k-1]}, x_k, \mathcal{V}) = p(x^{(-\infty, k-1]}, x_k, \mathcal{V}_0)$$

Gemäß Lemma (5.2) (vgl. die Bemerkung nach (5.6)) folgt daraus

durch Multiplikation

$$p(x^{-N}, x^{[0, n]}, \mathcal{V}) = p(x^{-N}, x^{[0, n]}, \mathcal{V}_0).$$

Die letzte Gleichung kann man für jedes $A \in \mathfrak{D} \in (-\infty, n]$ über

$A \cap \tilde{A}^n \in \mathfrak{D} \in (-\infty, n]$ integrieren.

$$\begin{aligned} & \int_{A \cap \tilde{A}^n} P(dx^{-N}) p(x^{-N}, x^{[0, n]}, \mathcal{V}) d1^{n+1} \\ & A \cap \tilde{A}^n \\ &= \int_{A \cap \tilde{A}^n} P(dx^{-N}) p(x^{-N}, x^{[0, n]}, \mathcal{V}_0) d1^{n+1} \end{aligned}$$

Es ist $\int_{A \cap \tilde{A}^n} = \int_A - \int_{A \cap \mathbb{C}\tilde{A}^n}$. Für das letzte der drei Integrale gilt:

$$\begin{aligned} 0 & \leq \int_{A \cap \mathbb{C}\tilde{A}^n} P(dx^{-N}) p(x^{-N}, x^{[0, n]}, \mathcal{V}_0) d1^{n+1} \\ & \leq \int_{\mathbb{C}\tilde{A}^n} P(dx^{-N}) p(x^{-N}, x^{[0, n]}, \mathcal{V}_0) d1^{n+1} \\ & = P(\mathbb{C}\tilde{A}^n) = 0 \end{aligned}$$

(5.10)c

Zufolge Lemma (7.8) ist

$$\{x^{(-\infty, n]} : p(x^{-N}, x^{[0, n]}, \mathcal{V}) = 0\}$$

nicht von $\mathcal{V}$ abhängig. Deshalb ist auch

$$\int_{A \cap \mathbb{C}\tilde{A}^n} P(dx^{-N}) p(x^{-N}, x^{[0, n]}, \mathcal{V}) d1^{n+1} = 0.$$

Damit ist gezeigt, daß für alle $A \in \mathfrak{D} \in (-\infty, n]$ gilt:

$$\int_A P(dx^{-N}) p(x^{-N}, x^{[0, n]}, \mathcal{V}) d1^{n+1} =$$

$$= \int_A P(dx^{-N}) P(x^{-N}, x^{[0,n]}, \mathcal{F}_0) d1^{n+1}$$

Auf Grund dieser Gleichung und wegen der Bemerkung nach

(5.6) gibt es nun für jedes $n \in \mathbb{N}$, $A^n \in \mathfrak{K}^n$ Elemente

$s := u(x^{-N}, \mathcal{F})$, $s_0 := u(x^{-N}, \mathcal{F}_0)$ aus S , so daß für alle $k \in \mathbb{N}$ gilt: $P^k(s, A^n, \mathcal{F}) = P^k(s_0, A^n, \mathcal{F}_0)$. Daraus folgt

$$\frac{1}{m} \sum_{k=1}^m P^k(s, A^n, \mathcal{F}) = \frac{1}{m} \sum_{k=1}^m P^k(s_0, A^n, \mathcal{F}_0),$$

$P(A^n, \mathcal{F}) = P(A^n, \mathcal{F}_0)$ gemäß Voraussetzung (7.2) und schließlich

$$P(\mathcal{F})| \mathfrak{K}^{\mathbb{Z}} = P(\mathcal{F}_0)| \mathfrak{K}^{\mathbb{Z}}.$$

Wir folgen nun dem durch Wald /27/ vorgezeichneten Weg, wie dies auch die Autoren von /2/ und /23/ taten.

(7.11) Definition Für $\delta > 0$ bezeichne

$$p^{(\delta)}(sx^{n-1}, x_n, \mathcal{F}) := \sup_{\mathcal{F}' \in U_\delta(\mathcal{F})} p(sx^{n-1}, x_n, \mathcal{F}')$$

$$p^{(\delta)}(x^{-N}, x_0, \mathcal{F}) := \sup_{\mathcal{F}' \in U_\delta(\mathcal{F})} p(x^{-N}, x_0, \mathcal{F}')$$

mit $U_\delta(\mathcal{F}) := \{ \mathcal{F}' \in \Theta : d(\mathcal{F}, \mathcal{F}') < \delta \}$, d Euklidische Metrik im $\mathbb{R}^m$

(7.12) Bemerkungen

1. Nach (D_1) ist $p(sx^{n-1}, x_n, \mathcal{F})$ stetig in $\mathcal{F} \in \Theta$ und nach

Lemma (7.5) ist auch $p(x^{-N}, x_0, \mathcal{F})$ stetig in $\mathcal{F} \in \Theta$ (jeweils für

festes $s \in S$, $x^n \in X^n$ bzw. $(x^{-N}, x_0) \in X^{(-\infty, 0]}$). Da es eine

abzählbare, dichte Menge $\{\mathcal{F}_i\} \subset \Theta$ gibt, kann man demnach

$\sup_{\mathcal{F}' \in U_\delta(\mathcal{F})}$ ersetzen durch $\sup_{i \in \mathbb{N}} p(x, \mathcal{F}_i)$, woraus folgt, daß die

in (7.11) definierten Funktionen $\mathfrak{K}^n$ bzw. $\mathfrak{K}^{(-\infty, 0]}$ meßbar sind.

2. Aus $(D_3)^*$ folgt zwar, daß

$$E(\log p^{(\delta)}(x^{-N}, x_0, \vartheta))^\alpha < \infty,$$

nicht aber, daß auch

$$E \log p^{(\delta)}(x^{-N}, x_0, \vartheta) < \infty.$$

Deswegen ersetzen wir $(D_3)^*$ durch

(D_3) Für alle $\vartheta \in \Theta$ gilt:

$$E \log p^{(\delta)}(x^{-N}, x_0, \vartheta) < \infty \text{ für ein } \delta > 0 \text{ und}$$

$$E(\log p^{(\delta)}(x^{-N}, x_0, \vartheta))^\alpha < \infty \text{ für ein } \alpha > 1.$$

(7.13) Lemma Es gilt für jedes $\vartheta \in \Theta$

$$\lim_{\delta \searrow 0} E \log p^{(\delta)}(x^{-N}, x_0, \vartheta) = E \log p(x^{-N}, x_0, \vartheta)$$

Beweis Wegen der Stetigkeit von $p(x^{-N}, x_0, \vartheta)$ in $\vartheta \in \Theta$ gilt
 $p^{(\delta)}(x^{-N}, x_0, \vartheta) \searrow p(x^{-N}, x_0, \vartheta)$ für $\delta \searrow 0$.

Wegen Voraussetzung (D_3) liefert nun der Satz von der monotonen Konvergenz die Behauptung.

Nun benötigen wir Satz (6.5) in einer Form, in der

$p^{(\delta)}(sx^{k-1}, x_k, \vartheta)$ an die Stelle von $p(sx^{k-1}, x_k)$ tritt.

(7.14) Satz Zu $s \in S$ und $\vartheta \in \Theta$ existiert ein $A \in \mathcal{X}^{\mathbb{N}}$ mit
 $p_s(A) = 1$, so daß $p^{(\delta)}(sx^{k-1}, x_k, \vartheta) > 0$ (für alle $k \in \mathbb{N}$) und

$$\lim_n \frac{1}{n} \sum_{k=1}^n \log p^{(\delta)}(sx^{k-1}, x_k, \vartheta) = E \log p^{(\delta)}(x^{-N}, x_0, \vartheta)$$

für alle $x^{\mathbb{N}} \in A$ gilt.

Beweis Der Beweis besteht zunächst darin, daß die zu Satz (6.5) führenden Lemmata (6.1) - (6.4) in geeigneter Weise neu formuliert werden.

Im folgenden wird $s_{\mathcal{J}}$ anstelle von s geschrieben, wenn ausgedrückt werden soll, daß s noch von $\mathcal{J}$ abhängen kann, wie es z.B. in $s_{\mathcal{J}} = u(s, x^m, \mathcal{J})$ der Fall ist. "Für alle $s_{\mathcal{J}}$ " heißt dann "für alle Abbildungen $\Theta \rightarrow S$ ", und $p^{(\delta)}(s_{\mathcal{J}} x^{n-1}, x_n, \mathcal{J})$ heißt $\sup_{\mathcal{J}' \in U_{\delta}(\mathcal{J})} p(s_{\mathcal{J}'} x^{n-1}, x_n, \mathcal{J}')$.

(6.1) gilt nun mit der Ungleichung

$$|p^{(\delta)}(s_{\mathcal{J}} x^{n-1}, x_n, \mathcal{J}) - p^{(\delta)}(\tilde{s}_{\mathcal{J}} x^{n-1}, x_n, \mathcal{J})| \leq C \beta^n$$

für alle $s_{\mathcal{J}}, \tilde{s}_{\mathcal{J}}$.

Das folgt aus (7.7) und der Ungleichung

$$|\sup_{\mathcal{J}} f(\mathcal{J}) - \sup_{\mathcal{J}} \tilde{f}(\mathcal{J})| \leq \sup_{\mathcal{J}} |f(\mathcal{J}) - \tilde{f}(\mathcal{J})|$$

(6.2) gilt mit der Ungleichung

$$p^{(\delta)}(x^{(-\infty, n-1]}, x_n, \mathcal{J}) \geq \varepsilon_n$$

wobei wieder, wie in (7.7), für die Folge ε_n gelten muß:

$$\sum_{n=1}^{\infty} (\log^{-} \varepsilon_n)^{-\alpha} < \infty.$$

Dies folgt aus (7.7) und der Ungleichung

$$p^{(\delta)}(x^{(-\infty, n-1]}, x_n, \mathcal{J}) \geq p(x^{(-\infty, n-1]}, x_n, \mathcal{J})$$

(6.3) gilt in der Form

- (a) $p^{(\delta)}(s_{\mathcal{J}} x^{n-1}, x_n, \mathcal{J}) > 0$ für alle $n \in \mathbb{N}$ und für alle $s_{\mathcal{J}}$
 (b) $|\log p^{(\delta)}(s_{\mathcal{J}} x^{n-1}, x_n, \mathcal{J}) - \log p^{(\delta)}(\tilde{s}_{\mathcal{J}} x^{n-1}, x_n, \mathcal{J})| \rightarrow 0$
 für alle $n \in \mathbb{N}, s_{\mathcal{J}}, \tilde{s}_{\mathcal{J}}$.

Dies folgt wieder aus (6.1) und (6.2).

(6.4) gilt in der Form

$$\lim_n \frac{1}{n} \sum_{k=1}^n \log p^{(\delta)}(x^{(-\infty, k-1]}, x_k, \mathcal{J}) = E \log p^{(\delta)}(x^{-N}, x_0, \mathcal{J}),$$

weil aufgrund von (D_3) das Ergodentheorem auf die Funktion

$p^{(\delta)}(x^{-N}, x_0, \mathcal{J})$ anwendbar ist.

Im Beweis zu (6.5) schließlich definiert man wieder

A_5 wie dort, wobei jetzt

$$(*) \quad p^{(\delta)}(s_{\mathcal{J}} x^{n-1}, x_n, \mathcal{J}) > 0 \quad \text{für alle } n \geq N \text{ und } s_{\mathcal{J}}$$

$$(**) \quad \lim_n \frac{1}{n} \sum_{k=N}^n \log p^{(\delta)}(s_{\mathcal{J}} x^{k-1}, x_k, \mathcal{J}) = E \log p^{(\delta)}(x^{-N}, x_0, \mathcal{J})$$

für alle $s_{\mathcal{J}}$

lauten. Man setzt A^* wie gehabt, und es gilt wieder (i) und (ii). Im Nachweis von (iii) heißt es dann hier :

"Nach Definition von A_5 existiert ein $N \in \mathbb{N}$ mit

$$p^{(\delta)}(\tilde{s}_{\mathcal{J}} x^{[m+1, m+n-1]}, x_{m+n}, \mathcal{J}) > 0 \quad \text{für alle } n \geq N \text{ und für alle } \tilde{s}_{\mathcal{J}} \text{ und}$$

$$\lim_n \frac{1}{n} \sum_{k=N}^n \log p^{(\delta)}(\tilde{s}_{\mathcal{J}} x^{[m+1, m+k-1]}, x_{m+k}, \mathcal{J})$$

$$= E \log p^{(\delta)}(x^{-N}, x_0, \mathcal{J}) \quad \text{für alle } \tilde{s}_{\mathcal{J}} .$$

Setze nun $\tilde{s}_{\mathcal{J}} := u(s_{\mathcal{J}}, x^m, \mathcal{J})$. Dann folgt wegen

$$u(\tilde{s}_{\mathcal{J}}, x^{[m+1, m+n]}, \mathcal{J}) = u(s_{\mathcal{J}}, x^{m+n}, \mathcal{J})$$

$$p^{(\delta)}(s_{\mathcal{J}} x^{n-1}, x_n, \mathcal{J}) > 0 \quad \text{für alle } n \geq N + m \text{ und für alle } s_{\mathcal{J}} \text{ und } \dots$$

$$\lim_n \frac{1}{n} \sum_{k=N+m}^n \log p^{(\delta)}(s_{\mathcal{J}} x^{k-1}, x_k, \mathcal{J}) = E \log p^{(\delta)}(x^{-N}, x_0, \mathcal{J})$$

für alle $s_{\mathcal{J}}$, womit Behauptung (iii) bewiesen ist".

Der Rest überträgt sich ohne Schwierigkeit mit Hilfe von (D_2) .

Nun können wir einen Satz über die (starke) Konsistenz der Maximum-Likelihood-Schätzung beweisen:

(7.15) Satz Vorausgesetzt werden die gleichmäßige Ergodizität ,

$(B_2) - (B_4)$, $(C_1) - (C_2)$, jeweils für jedes $\mathcal{J} \in \Theta$, und

$(D_1) - (D_3)$. Dann gibt es zu jedem $s \in S$ ein $A \in \mathfrak{A}^{\mathbb{N}}$ mit

$P_s(A) = 1$, so daß für alle $x^{\mathbb{N}} \in A$ gilt:

Existiert für jedes $n \in \mathbb{N}$ ein absolutes Maximum von $p(s, x^n, \vartheta)$ über Θ an der (nicht notwendig eindeutigen) Stelle

$\hat{\vartheta}_n = \hat{\vartheta}_n(x^n)$, so gilt für jeden Häufungspunkt $\hat{\vartheta}$ der Folge $(\hat{\vartheta}_n)_{n \in \mathbb{N}}$:

$$P(\hat{\vartheta}) = P(\vartheta_0)$$

Beweis Sei $\delta > 0$ und $U_\delta(\hat{\vartheta}) := \{ \vartheta \in \Theta : d(\vartheta, \hat{\vartheta}) < \delta \}$.

Es gilt $\hat{\vartheta}_n \in U_\delta(\hat{\vartheta})$ für unendlich viele $n \in \mathbb{N}$ und damit auch

$$\frac{1}{n} \log p(s, x^n, \vartheta_0) \leq \frac{1}{n} \log p(s, x^n, \hat{\vartheta}_n)$$

$$= \sup_{\vartheta' \in U_\delta(\hat{\vartheta})} \frac{1}{n} \log p(s, x^n, \vartheta')$$

$$\leq \frac{1}{n} \sum_{k=1}^n \sup_{\vartheta' \in U_\delta(\hat{\vartheta})} \log p(s x^{k-1}, x_k, \vartheta')$$

$$= \frac{1}{n} \sum_{k=1}^n \log p^{(\delta)}(s x^{k-1}, x_k, \hat{\vartheta}) \quad \text{für unendlich viele Indizes } n \in \mathbb{N}$$

Gemäß (6.5) und (7.14) folgt daraus für P_S -fast alle $x^{\mathbb{N}} \in X^{\mathbb{N}}$

$$h(\vartheta_0) \leq E \log p^{(\delta)}(x^{\mathbb{N}}, x_0, \hat{\vartheta}) \quad (\text{das } x^{\mathbb{N}} \in X^{\mathbb{N}} \text{ geht in das } \hat{\vartheta} \text{ ein.})$$

Gemäß (7.13) gilt dann

$$h(\vartheta_0) \leq h(\hat{\vartheta}),$$

so daß Satz (7.10) die Behauptung liefert.

(7.16) Bemerkung

1. Ist $\Theta \subset \mathbb{R}^m$ kompakt, dann existiert (für festes s, n, x^n) (mindestens) ein absolutes Maximum von $p(s, x^n, \vartheta)$ und (mindestens) ein Häufungspunkt $\hat{\vartheta}$ von $(\hat{\vartheta}_n)_{n \in \mathbb{N}}$.

Gilt darüber hinaus auch $P(\vartheta) \neq P(\vartheta_0)$ für $\vartheta \neq \vartheta_0$, dann findet für P_S -fast alle $x^{\mathbb{N}}$ Konvergenz $\hat{\vartheta}_n \rightarrow \vartheta_0$ statt.

2. Im allgemeinen folgt aus $P(\vartheta) = P(\vartheta_0)$ nicht $P_S(\vartheta) = P_S(\vartheta_0)$

Als Beispiel nehme man den Markovprozeß, der zu der

stochastischen $N \times N$ Matrix $\bigwedge$ in Definition (4.3) (*) gehört, und $\mathcal{J} := \bigwedge \in \mathbb{R}^{N^2}$. Es ist dann $P(\mathcal{J}') = P(\mathcal{J}'')$ genau dann, wenn die zu $\bigwedge'$ bzw. $\bigwedge''$ gehörenden R' und R'' gleich sind. Es lassen sich also in diesem Fall für diejenigen $\mathcal{J}$ -Komponenten, die zu "transienten" Ereignissen gehören, keine konsistenten Schätzungen finden.

3. Die in Satz (7.15) auftretende Maximum-Likelihood-Funktion $\hat{\mathcal{J}}_n(x^n)$ ist für lokalkompaktes Θ nach /21/, Theorem 3.10, $\mathcal{K}^n$ -meßbar für jedes $n \in \mathbb{N}$.

4. Es bleibt das Problem offen, unter welchen zusätzlichen Voraussetzungen sich wie in /5/ und /24/ die asymptotische Normalverteilung von $\sqrt{n} (\hat{\mathcal{J}}_n - \mathcal{J}_0)$ und die asymptotische χ^2 -Verteilung von $2(\log p(s, x^n, \hat{\mathcal{J}}_n) - \log p(s, x^n, \mathcal{J}_0))$ zeigen läßt. Dies würde uns den Schlüssel zu vielen praktischen Testverfahren liefern. (vgl. /5/, Theorem 2.2 und Paragraphen 3 und 4, und /24/)

5. Der zweite Teil der Bedingung (D_3) kann bei unveränderter Beweisführung so verallgemeinert werden (vgl. Lemma (7.7)):

(*) $E (\log^{-\alpha} p(x^{-N}, x_0, \vartheta))^\alpha < \infty$ mit einem solchen $\alpha > 0$,

daß für P-fast alle x^N für die Folge (vgl. Lemma (6.1))

$\varepsilon_n := \sup_{s', s'' \in S, \vartheta \in \Theta} |p(s' x^{n-1}, x_n, \vartheta) - p(s'' x^{n-1}, x_n, \vartheta)|$, $\varepsilon_n < 1$ für $n \geq N(x^N)$, gilt: $\sum_{n \geq N} (\log^{-\alpha} \varepsilon_n)^{-\alpha} < \infty$. Gilt $\varepsilon_n = 0$ ab einem gewissen $N(x^N)$,

so kann man in (*) formal $\alpha = 0$ setzen, d.h., auf die Bedingung

(*) ganz verzichten. Für $\varepsilon_n = \beta^{(n^\gamma)}$, $\beta < 1, \gamma > 0$, z.B. muß $\alpha > 1/\gamma$

gesetzt werden, für einen Markovprozess endlicher Ordnung z.B.

kann man auf Bedingung (*) verzichten.

8. Beispiele

In diesem Paragraphen sollen Beispiele angeführt werden, auf welche der Satz (7.15) von der Konsistenz der Maximum-Likelihood-Schätzung anwendbar ist. Anwendbar ist dieser Satz jedenfalls dann, wenn die Bedingungen:

gleichmäßige Ergodizität, $(B_2) - (B_4)$, (C_1) , (C_2) , jeweils für alle $\vartheta \in \Theta$, und $(D_1) - (D_3)$ erfüllt sind.

(8.1) Lineare OM-Kette (Definition (2.6), Satz (2.9))

Die $\alpha_x(\vartheta)$, $\bigwedge_x(\vartheta)$ mögen von $\vartheta \in \Theta \subset \mathbb{R}^m$ abhängen, so daß $0 \leq \alpha_x(\vartheta) \leq \lambda$ ($\lambda < 1$) gilt. Lemma (3.2) zeigt, daß dann $(B_2) - (B_4)$ für jedes $\vartheta \in \Theta$ gilt. Die Voraussetzungen von Satz (5.12)_b, sind sie für jedes $\vartheta \in \Theta$ erfüllt, liefern dann (C_1) , (C_2) für jedes $\vartheta \in \Theta$.

Die 1-Dichte $p(\bigwedge_x(\vartheta), y)$ von $\bigwedge_x(\vartheta)$ soll mit $f(x, y, \vartheta)$ abgekürzt werden. Sind $\alpha_x(\vartheta)$ und $f(x, y, \vartheta)$ stetig in $\vartheta \in \Theta$ für jedes $x \in X$, so folgt aus (*) im Beweis zu (5.12) sofort die Gültigkeit von (D_1) .

(8.2) Lemma Die Mengen $\{x \in X : \alpha_x(\vartheta) = 0\}$ und $\{(x, y) \in X^2 : f(x, y, \vartheta) = 0\}$ seien von ϑ unabhängig. Dann gilt (D_2) .

Beweis folgt aus (*) im Beweis zu (5.12) mit Hilfe der gemachten Voraussetzungen.

Es bleibt das Problem offen, Voraussetzungen für eine lineare OM-Kette zu finden, aus denen Bedingung (D_3) ableitbar ist.

(8.3) Markovprozeß

Wir betrachten nun einen Markovprozeß mit Zustandsraum X .

Der zugehörige Markovkern $\bigwedge_x(A, \vartheta)$, $x \in X$, $A \in \mathcal{X}$, $\vartheta \in \Theta$ besitze die 1-Dichte $f(x, y, \vartheta)$; $f^{(\delta)}(x, y, \vartheta) := \sup_{\vartheta' \in U_\delta(\vartheta)} f(x, y, \vartheta')$

Für jedes $\vartheta \in \Theta$ gelte

(M₁) der Markovprozeß ist stationär und ergodisch

(M₂) $f(x, y, \vartheta)$ ist $\mathcal{X} \times \mathcal{X}$ -meßbar

(M₃) $E \log f^{(\delta)}(x, y, \vartheta) < \infty$ für ein $\delta > 0$

Weiterhin gelte

(M₄) $f(x, y, \vartheta)$ ist für jedes $(x, y) \in X^2$ stetig in $\vartheta \in \Theta$

(M₅) $\{(x, y) \in X^2 : f(x, y, \vartheta) = 0\}$ ist nicht von ϑ abhängig.

Dann ist Satz (7.15) anwendbar.

Beweis durch die Spezialisierung $\alpha_x = 0$ für alle $x \in X$ in der linearen OM-Kette unter Beachtung der Bemerkung (7.16) 5.

(8.4) Bemerkung Die Voraussetzungen (M₁) - (M₅) sind im wesentlichen äquivalent mit den Voraussetzungen (A₁) - (A₈) bei Roussas /23/ oder den Voraussetzungen (2.5) - (2.8) bei Gänssler /11/. Allein (M₄) findet sich dort in dieser abgeschwächten Form: $f(x, y, \vartheta)$ ist nach oben halbstetig in ϑ und $f^{(\delta)}(x, y, \vartheta)$ meßbar in (x, y) . (Bei Stetigkeit anstelle der Halbstetigkeit folgt die Meßbarkeit von $f^{(\delta)}(x, y, \vartheta)$ schon aus der Separabilität von $\Theta \subset \mathbb{R}^m$). Eine solche Abschwächung läßt aber unsere Voraussetzung (M₄) bzw. (D₁) ebenfalls ohne weiteres zu. Gänssler erarbeitet im übrigen noch Voraussetzungen, die von der speziellen Version $f(x, y, \vartheta)$ der Dichte von $\bigwedge_x(A, \vartheta)$ unabhängig sind.

Im Falle ergodischer Markovprozesse ist das Maximum-Likelihood-Schätzproblem gegenüber unserer Situation insofern einfacher gelagert, als man stets voraussetzen kann, daß die Startverteilung gleich der stationären Verteilung ist, so daß das Problem der Nichtstationarität von Prozessen gar nicht auftaucht.

Nun behandeln wir den Fall der endlichen Markovkette, wo wir den Parameter ϑ identisch der Übergangsmatrix Λ setzen.

(8.5) Endliche Markovkette

Es sei $X := \{1, 2, \dots, N\}$, Θ eine Teilmenge der Menge aller stochastischen $N \times N$ Matrizen Λ , s der Startvektor der Markovkette. Dann wird

$$\begin{aligned} \frac{1}{n} \log P(s, x^n, \Lambda) &= \frac{1}{n} \log s_{x_1} \Lambda_{x_1 x_2} \dots \Lambda_{x_{n-1} x_n} \\ &= \frac{1}{n} \log s_{x_1} + \frac{1}{n} \sum_{i,j=1}^N \nu_{ij}(x^n) \log \Lambda_{ij} \end{aligned}$$

wobei $\nu_{ij}(x^n)$ die Anzahl der k , $1 \leq k \leq n-1$, mit $x_k = i, x_{k+1} = j$ bedeutet. Das eindeutige absolute Maximum von $\log P(s, x^n, \Lambda)$ über alle stochastischen $N \times N$ Matrizen Λ wird nun an der Stelle

$$(*) \quad \hat{\Lambda}_{ij}(x^n) = \nu_{ij}(x^n) / \sum_{j=1}^N \nu_{ij}(x^n)$$

angenommen. Von den Bedingungen $(M_1) - (M_5)$ bleiben hier nur noch (M_1) und (M_5) übrig, wobei (M_5) jetzt lautet:

Die Menge $\{(i, j) \in X^2 : \hat{\Lambda}_{ij} = 0\}$ ist von $\Lambda \in \Theta$ unabhängig.

(8.3) liefert nun das folgende Korollar, das sich in äquivalenter Form zum Beispiel auch in /5/, Cond. 5.1/Theorem 5.3 findet.

(8.6) Korollar Der Parameterraum Θ einer endlichen Markovkette erfülle (M_1) und (M_5) . Dann ist die in (8.5) (*) angegebene Maximum-Likelihood-Schätzung konsistent.

(8.7) Endliche lineare OM-Kette (vgl. (4.1))

Hier ist $\Theta = (\Theta_\alpha, \Theta_\wedge)$ mit $\Theta_\alpha \subset [0,1]^N$ und $\Theta_\wedge$ Teilmenge der Menge aller stochastischen $N \times N$ Matrizen. Wir formulieren:

(E₁) Jedes $\wedge \in \Theta_\wedge$ ist ergodisch (im Sinne von (4.3)).

(E₂) Die Mengen $\{i \in X : \alpha_i = 0\}$ und $\{(i,j) \in X^2 : \wedge_{ij} = 0\}$ sind unabhängig von $\alpha \in \Theta_\alpha$ und $\wedge \in \Theta_\wedge$

(E₃) Mit einem $\lambda < 1$ gilt:

$$\alpha_i \leq \lambda \quad \text{für alle } \alpha \in \Theta_\alpha, i \in X.$$

(8.8) Satz Der Parameterraum $(\Theta_\alpha, \Theta_\wedge)$ einer endlichen linearen OM-Kette erfülle (E₁) - (E₃) .

Dann ist Satz (7.15) anwendbar.

Beweis Aufgrund von Satz (4.5) folgt aus (E₁)₁ und (E₃)₃ daß die Kette mvB für jedes $\vartheta \in \Theta$ gleichmäßig ergodisch ist. Durch Spezialisierung von (8.1) und (8.2) folgt aus (E₂) und (E₃) die Gültigkeit von (B₂) - (B₄) für alle $\vartheta \in \Theta$, (C₁) , (C₂) für alle $\vartheta \in \Theta$ und (D₁) , (D₂) . Es bleibt, die Gültigkeit von (D₃) zu zeigen. Ohne Einschränkung kann dafür $P(i) > 0$ für jedes $i \in X$ vorausgesetzt werden. Aus Lemma (2.11) schließt man

$$p(x^{-N}, x_0, \vartheta) = \sum_{i=1}^N \beta_i(x^{-N}, \vartheta) \wedge_{ix_0} \text{ mit}$$

$$\beta_i(x^{-N}, \vartheta) = (1-\alpha_i) \{ \alpha_{x_{-1}} \dots \alpha_{x_{-n_1}} + \alpha_{x_{-1}} \dots \alpha_{x_{-n_2}} + \dots \},$$

wobei die von i abhängenden $0 < n_1 < n_2 < \dots$ definiert sind durch

$x_{-(n_1+1)} = i$, $x_{-(n_2+1)} = i$, ... Die Abhängigkeit der $n_1, n_2, \dots$ von i soll in der Notation unterdrückt werden. Zur Abkürzung setzen wir weiter β_i anstelle von $\beta_i(x^{-N}, \vartheta)$ und β_i° anstelle von

$\beta_i(x^{-N}, \vartheta_0)$, wobei $\vartheta_0 = (\alpha^0, \wedge^0)$ der wahre Parameter ist.

Da 1 hier die Gleichverteilung auf X ist, gilt

$$E(\log p(x^{-N}, x_0, \vartheta))^\alpha \stackrel{=}{=} (5.10) \quad d$$

$$\begin{aligned} &= \int_{X(-\infty, 0]} p(x^{-N}, x_0, \vartheta_0) (\log p(x^{-N}, x_0, \vartheta))^\alpha P(dx^{-N}) \cdot 1(dx_0) \\ &= c_1 \sum_{X_0} \int_{X^{-N}} \left(\sum_{i=1}^N \beta_i^0 \wedge_{iX_0} \right) (\log \left(\sum_{i=1}^N \beta_i \wedge_{iX_0} \right))^\alpha P(dx^{-N}) \quad (c_1 := \frac{1}{N}) \\ &\leq c_1 \sum_{X_0} \int_{X^{-N}} \sum_{i=1}^N \beta_i^0 \wedge_{iX_0} (\log \beta_i \wedge_{iX_0})^\alpha P(dx^{-N}) \\ &= c_1 \sum_{i=1}^N \int_{X^{-N}} \sum_{X_0} \beta_i^0 \wedge_{iX_0} (\log \beta_i \wedge_{iX_0})^\alpha P(dx^{-N}) \end{aligned}$$

Für jedes i teilen wir nun X^{-N} disjunkt auf nach der Stelle, an welcher in der Folge $(x_{-1}, x_{-2}, \dots) \in X^{-N}$ zum erstenmal i auftritt. Alle x^{-N} , in denen keine Komponente gleich i ist, können wir gemäß Satz (3.19) vernachlässigen. Genauer setzen wir:

$$X^{-N} = \sum_{n_1=0}^{\infty} X_{n_1}^{(i)} \text{ mit } \{x^{-N}: x_{-1} \neq i, \dots, x_{-n_1} \neq i, x_{-(n_1+1)} = i\} =: X_{n_1}^{(i)}$$

Dann gilt für alle $x^{-N} \in X_{n_1}^{(i)}$

$$\alpha_{x_{-1}} \dots \alpha_{x_{-n_1}} + \alpha_{x_{-1}} \dots \alpha_{x_{-n_2}} + \dots = \alpha_{x_{-1}} \dots \alpha_{x_{-n_1}} (1 + B(x^{-N}, \vartheta))$$

mit

$$0 \leq B(x^{-N}, \vartheta) \leq \sum_{k=1}^{\infty} \lambda^k = \frac{\lambda}{1-\lambda} < \infty.$$

Damit können wir die obige Abschätzung fortsetzen

$$(c_2 := c_1 \frac{1}{1-\lambda}, \quad A_{ij} := (1-\alpha_i) \wedge_{ij}, \quad n := n_1)$$

$$\leq c_2 \sum_{i=1}^N \sum_{n=1}^{\infty} \int_{X_n^{(i)}} \sum_{X_0} [(\alpha_{x_{-1}}^0 \dots \alpha_{x_{-n}}^0 A_{iX_0}^0)]^{1/\alpha}.$$

$$\log (\alpha_{x_{-1}} \dots \alpha_{x_{-n}} A_{iX_0})]^\alpha P(dx^{-N})$$

$$\leq c_2 \sum_{i=1}^N \sum_{n=1}^{\infty} \int_{X_n^{(i)}} \sum_{X_0} \left[\sum_{j=1}^n (\alpha_{x_{-1}}^0 \dots \alpha_{x_{-n}}^0 A_{ix_0}^0)^{1/\alpha} \cdot \log^{-\alpha_{x_{-j}}} A_{ix_0} \right]^{\alpha} P(dx^{-N})$$

Aus der Voraussetzung (E_2) folgt für alle $i, j, k \in X$ mit einem $L' < \infty$

$$\alpha_j^0 A_{ik}^0 \log^{-\alpha_j} A_{ik} \leq L' \quad \text{und deshalb auch mit einem } L < \infty$$

$$(\alpha_{x_{-1}}^0 \dots \alpha_{x_{-n}}^0 A_{ix_0}^0)^{1/\alpha} \log^{-\alpha_{x_{-j}}} A_{ix_0} \leq L$$

Mit $c_3 := c_2 N L^{\alpha}$ setzen wir die Abschätzung fort:

$$\leq c_3 \sum_{i=1}^N \sum_{n=1}^{\infty} n^{\alpha} \int_{X_n^{(i)}} P(dx^{-N}) = c_3 \sum_{i=1}^N \sum_{n=1}^{\infty} n^{\alpha} P(X_n^{(i)})$$

Setzt man nun in Satz (3.19) $A := X - \{i\}$, so bekommt man

$$P(X_n^{(i)}) \leq P(\underbrace{(A, \dots, A)}_n) \leq c_4 d^n \quad \text{mit } c_4 < \infty, d < 1.$$

Wir erhalten also schließlich

$$E(\log^{-\alpha} P(x^{-N}, x_0, j))^{\alpha} \leq c_3 \cdot c_4 \cdot N \sum_{n=1}^{\infty} n^{\alpha} d^n < \infty \quad \text{w.z.b.w.}$$

9. Ein Iterationsverfahren zur Bestimmung der Maximum-Likelihood-Schätzung in linearen OM-Ketten

In diesem Paragraphen geht es um eine konstruktive Methode, in einer linearen OM-Kette die Maximumstelle $\hat{\vartheta}_n$ von $p(s, x^n, \vartheta)$ über Θ zu bestimmen. Das hier vorgestellte Verfahren lehnt sich an ein Verfahren an, das für sogenannte "probabilistic functions of Markov chains" entwickelt wurde /3/. Es galt für uns, zur Likelihoodfunktion $L(\vartheta) = p(s, x^n, \vartheta)$ eine geeignete Funktion $Q(\vartheta, \check{\vartheta})$ zu finden (die ein etwas anderes Aussehen hat als in /3/) so daß Satz (9.4), der sich in analoger Form in /3/ findet, gültig ist. Zusätzlich zu /3/ wird auch noch eine Nebenbedingung $g(\vartheta) = 0$ für $\vartheta \in \Theta$ eingeführt.

Im folgenden seien nun Anfangszustand μ und n, x^n fest gewählt; Θ wird als eine offene Teilmenge des $\mathbb{R}^m$ vorausgesetzt. Die Abbildung

$g : \Theta \subset \mathbb{R}^m \longrightarrow \mathbb{R}^r$ ($r < m$), $g(\vartheta) = (g_1(\vartheta), \dots, g_r(\vartheta))$, sei stetig differenzierbar.

(9.1) Erläuterungen Sei

$$L(\vartheta) := p(\mu, x^n, \vartheta) = \prod_{k=1}^n p(\mu x^{k-1}, x_k, \vartheta) \quad (5.2)$$

die Likelihoodfunktion in Abhängigkeit von $\vartheta \in \Theta$. Gemäß der Gleichung (*) im Beweis zu (5.12) gilt

$$p(\mu x^{k-1}, x_k, \vartheta) = \sum_{l=0}^{k-1} \beta(l, k-1, \vartheta) \cdot f(x_1, x_k, \vartheta), \text{ wobei}$$

$$\beta(l, m, \vartheta) := (1 - \alpha_{x_1}(\vartheta)) \alpha_{x_{l+1}}(\vartheta) \dots \alpha_{x_m}(\vartheta) \quad \text{und}$$

$$f(x, y, \vartheta) \text{ die } W\text{-Dichte von } \bigwedge_x(\cdot, \vartheta) \text{ ist.}$$

(Der Fall $l = 0$ versteht sich so:

$$\beta(0, m, \vartheta) := \alpha_{x_1}(\vartheta) \dots \alpha_{x_m}(\vartheta) \quad \text{und}$$

$$f(x_0, x, \vartheta) := W\text{-Dichte zum Anfangszustand } \mu .)$$

Setze nun zur Abkürzung

$$p_k(\vartheta) := p(\mu x^{k-1}, x_k, \vartheta)$$

$$p_{k1}(\vartheta) := \beta(1, k-1, \vartheta) f(x_1, x_k, \vartheta)$$

Dann ist

$$(*) \begin{cases} L(\vartheta) = \prod_{k=1}^n p_k(\vartheta) \\ p_k(\vartheta) = \sum_{l=0}^{k-1} p_{kl}(\vartheta) \end{cases} \quad \text{und}$$

(9.2) Definition Zu einer Likelihoodfunktion $L(\vartheta)$, die sich in der Form (9.1) (*) darstellen läßt, definieren wir die Funktion

$$Q(\vartheta, \tilde{\vartheta}) := \sum_{k=1}^n \frac{1}{p_k(\vartheta)} \sum_{l=0}^{k-1} p_{kl}(\vartheta) \log p_{kl}(\tilde{\vartheta}) \quad , \quad \vartheta, \tilde{\vartheta} \in \Theta .$$

Für das Folgende benötigen wir diese Voraussetzung

(D₄) Die Menge $\{(k, l) : p_{kl}(\vartheta) = 0\}$ ist unabhängig von ϑ .

Die $p_{kl}(\vartheta)$ sind stetig differenzierbar nach ϑ . Bei festem ϑ hat $Q(\vartheta, \tilde{\vartheta})$ als Funktion von $\tilde{\vartheta} \in \Theta$ ein eindeutiges Maximum unter der Nebenbedingung $g(\tilde{\vartheta}) = 0$. Die Stelle, an welcher dieses Maximum angenommen wird, ist auch eine kritische Stelle von $Q(\vartheta, \tilde{\vartheta})$ unter der Nebenbedingung $g(\tilde{\vartheta}) = 0$.

(Wir sagen, $\hat{\vartheta}$ ist eine kritische Stelle von $f(\vartheta)$ unter der Nebenbedingung $g(\vartheta) = 0$, wenn $g(\hat{\vartheta}) = 0$ und es ein $\lambda \in \mathbb{R}^r$ gibt mit $\text{grad } f|_{\hat{\vartheta}} + \lambda \cdot \text{grad } g|_{\hat{\vartheta}} = 0$. $\lambda \in \mathbb{R}^r$ heißt auch r -dimensionaler "Lagrange-Multiplikator").

(9.3) Definition Unter der Voraussetzung (D₄) bilden wir die Abbildung $T : \Theta \longrightarrow \Theta$ gemäß:

$T(\vartheta)$ ist eine (eindeutige) Maximumstelle von $Q(\vartheta, \tilde{\vartheta})$ über $\tilde{\vartheta} \in \Theta$ unter der Nebenbedingung $g(\tilde{\vartheta}) = 0$:

$$Q(\vartheta, T(\vartheta)) = \max_{\tilde{\vartheta} \in \Theta, g(\tilde{\vartheta}) = 0} Q(\vartheta, \tilde{\vartheta}) .$$

Für die Abbildung T gilt nun der folgende Satz, der sich in /3/ unter Theorem 2.1, Proposition 2.1, Theorem 3.1 findet:

(9.4) Satz Voraussetzung (D_4) sei erfüllt. Dann gilt $L(T(\vartheta)) > L(\vartheta)$, außer wenn $T(\vartheta) = \vartheta$. Sei $\vartheta_1 \in \Theta$. Jeder Häufungspunkt der Folge $(T^n(\vartheta_1))_{n \in \mathbb{N}}$, wobei T^n die n -te Iterierte von T ist, ist kritische Stelle von $L(\vartheta)$ unter der Nebenbedingung $g(\vartheta) = 0$.

Beweis Zuerst zeigt man analog zum Beweis von Theorem 2.1 in /3/, daß $L(\tilde{\vartheta}) \geq L(\vartheta)$, falls $Q(\vartheta, \tilde{\vartheta}) \geq Q(\vartheta, \vartheta)$:

$$\begin{aligned} 0 &\leq Q(\vartheta, \tilde{\vartheta}) - Q(\vartheta, \vartheta) \\ &= \sum_k \sum_l \frac{p_{kl}(\vartheta)}{p_k(\vartheta)} \log \frac{p_{kl}(\tilde{\vartheta})}{p_{kl}(\vartheta)} \\ &\leq \sum_k \log \sum_l \frac{p_{kl}(\vartheta)}{p_k(\vartheta)} \frac{p_{kl}(\tilde{\vartheta})}{p_{kl}(\vartheta)} \\ &= \log \prod_k \frac{p_k(\tilde{\vartheta})}{p_k(\vartheta)} = \log(L(\tilde{\vartheta})/L(\vartheta)) \end{aligned}$$

also $L(\tilde{\vartheta}) \geq L(\vartheta)$, wobei

$$\sum_l \frac{p_{kl}(\vartheta)}{p_k(\vartheta)} = 1 \text{ und die Jensensche Ungleichung ausgenutzt wurden.}$$

Weiter rechnet man nach, daß

$$\left. \frac{\partial Q(\vartheta, \tilde{\vartheta})}{\partial \tilde{\vartheta}_1} \right|_{\tilde{\vartheta}=\vartheta} = \left. \frac{\partial \log L(\vartheta)}{\partial \vartheta_1} \right|_{\vartheta} , \text{ woraus folgt, daß jeder}$$

Fixpunkt von T eine kritische Stelle von $L(\vartheta)$ unter der Nebenbedingung $g(\vartheta) = 0$ ist. Gilt $\tilde{\vartheta} := T(\vartheta) \neq \vartheta$, so ist nach

Voraussetzung (D_4) $Q(\vartheta, \tilde{\vartheta}) > Q(\vartheta, \vartheta)$, woraus nach der obigen Abschätzung auch $L(\tilde{\vartheta}) > L(\vartheta)$ folgt.

Wie in /3/, Proposition 2.1, zeigt man schließlich, daß jeder Häufungspunkt der Folge $(T^n(\vartheta_1))$ ein Fixpunkt von T ist und folglich die im Satz behauptete Eigenschaft besitzt.

(9.5) Bemerkung Satz (9.4) beinhaltet die Vorschrift für ein Iterationsverfahren zur Berechnung der Maximumstelle $\hat{\vartheta}$ von $L(\vartheta)$ unter der Nebenbedingung $g(\vartheta) = 0$ (sofern eine solche Stelle überhaupt existiert). Man beachte, daß das Verfahren nur dann $\hat{\vartheta}$ liefert, wenn $\hat{\vartheta}$ auch kritische Stelle (d.h. auch "lokales" Maximum) von $L(\vartheta)$ unter der Nebenbedingung $g(\vartheta) = 0$ ist, und, falls es noch weitere solcher kritischen Stellen gibt, ϑ_i genügend nahe bei $\hat{\vartheta}$ gewählt wird.

Wir wenden uns nun der endlichen linearen OM-Kette mit $X = \{1, 2, \dots, N\}$ und $\vartheta = (\alpha, \wedge)$ zu.

(9.6) Abkürzungen (vgl. auch (9.1); es wird ab jetzt die ϑ -Abhängigkeit in der Notation unterdrückt.)

$$\beta(l, m) := (1 - \alpha_{x_1}) \alpha_{x_{1+1}} \dots \alpha_{x_m} \quad (l \leq m)$$

$$\gamma(i, m) := \sum_{1 \leq l_1 \leq m} \beta(l_1, m) \quad \text{für jedes } i \in X,$$

wobei l_1 diejenigen Stellen in der Sequenz $(x_1, \dots, x_m)$ durchläuft, an denen ein i auftritt ($x_{l_1} = i$).

$$p_k := \sum_{l=0}^{k-1} \beta(l, k-1) \wedge x_1 x_k$$

$$(\beta(0, m) := \alpha_{x_1} \dots \alpha_{x_m}, \wedge_{x_0 x} := s_x, s \text{ Startvektor} \in \mathbb{R}^N)$$

$$\beta^{(i)}(1,m) := \beta(1,m) \cdot n_i[1+1,m] ,$$

wobei $n_i[1+1,m]$ die Anzahl der vorkommenden i in der Sequenz $(x_{1+1}, \dots, x_m)$ bezeichnet,

$$\gamma^{(i)}(j,m) := \sum_{1 \leq l_j \leq m} \beta^{(i)}(1_j, m)$$

(9.7) Satz Für die endliche lineare OM-Kette ist (D_4) erfüllt, wenn $\alpha_i < 1$ für alle $i \in X$ und (E_2) aus Paragraph 8 gilt. Die in (9.3) definierte Abbildung T lautet (Setze $(\tilde{\alpha}, \tilde{\wedge}) := T(\alpha, \wedge)$)

$$\tilde{\alpha}_i := \frac{\sum_{k=1}^n \left\{ \sum_{j=0}^N \gamma^{(i)}(j, k-1) \wedge_{j x_k} \right\} / p_k}{\sum_{k=1}^n \left\{ \gamma(i, k-1) \wedge_{i x_k} + \sum_{j=0}^N \gamma^{(i)}(j, k-1) \wedge_{j x_k} \right\} / p_k}$$

$$(\gamma^{(i)}(0, m) := \beta^{(i)}(0, m) , \wedge_{0 x_k} := s_{x_k} , \text{ vgl. (9.6))}$$

$$\tilde{\wedge}_{ij} := \frac{\sum_{k=1}^n \left\{ \gamma(i, k_j-1) \wedge_{ij} \right\} / p_{k_j}}{\sum_{k=1}^n \left\{ \gamma(i, k-1) \wedge_{i x_k} \right\} / p_k}$$

Beweis Es ist

$$p_{k1} = (1 - \alpha_{x_1}) \alpha_{x_{1+1}} \dots \alpha_{x_{k-1}} \wedge_{x_1 x_k} . \text{ Wegen } \alpha_i < 1 \text{ und } (E_2) \text{ ist die}$$

erste Forderung in (D_4) erfüllt, ebenso wie die zweite.

Es ist $\log p_{k1}$ streng konkav in $(\alpha, \wedge)$ und damit auch $Q(\varphi, \tilde{\varphi})$

für jedes φ streng konkav in $\tilde{\varphi} = (\tilde{\alpha}, \tilde{\wedge})$. (D_4) ist erfüllt,

wenn wir noch zeigen können, daß $Q(\varphi, \tilde{\varphi})$ als Funktion von $\tilde{\varphi}$

einen kritischen Punkt unter der linearen Nebenbedingung

$\sum_{j=1}^N \tilde{\wedge}_{ij} - 1 = 0$, $i = 1, 2, \dots, N$ hat. Man rechnet nun nach, daß das Gleichungssystem

$$\frac{\partial Q(\tilde{\nu}, \tilde{\nu})}{\partial \tilde{\Lambda}_{ij}} + \lambda_i = 0 \quad i, j = 1, \dots, N, \lambda_i \text{ Langrange-Multiplikator}$$

$$\frac{\partial Q(\tilde{\nu}, \tilde{\nu})}{\partial \tilde{\alpha}_i} = 0 \quad i = 1, \dots, N$$

$$\sum_{j'} \tilde{\Lambda}_{ij'} - 1 = 0 \quad i = 1, \dots, N$$

die oben angegebene Lösung $(\tilde{\alpha}, \tilde{\Lambda})$ besitzt, womit der Satz bewiesen ist.

(9.8) Bemerkungen

1. Setzt man $\alpha_i = 0$ für alle $i \in X$, so erhält man $\tilde{\alpha}_i = 0$ und $\tilde{\Lambda}_{ij} = \hat{\Lambda}_{ij}$ (vgl. (8.5) (*)).

D.h. die Schätzung $\hat{\Lambda}$ der Markovübergangsmatrix ist in (9.7) als Sonderfall enthalten.

2. Von der Erweiterung, welche die Schätzung $\tilde{\Lambda}$ in (9.7) gegenüber der Schätzung $\hat{\Lambda}$ in (8.5) darstellt, erhält man eine Vorstellung, wenn man in (9.7) die Näherung $\bigwedge_{i, k_k} / p_k = 1$ macht, d.h., wenn man in (9.7) die Abhängigkeit der $\tilde{\Lambda}_{ij}$ von $\bigwedge$ eliminiert. Setzt man

$$\gamma_{ij} := \sum_{1 \leq k_j \leq n} \beta(i, k_j - 1) \stackrel{(9.6)}{=} \sum_{\substack{1 \leq k_j \leq n \\ 1 \leq l_1 \leq k_j - 1}} \beta(l_1, k_j - 1),$$

so erhält man

$$\tilde{\Lambda}_{ij} = \frac{\gamma_{ij}}{\sum_{j'} \gamma_{ij'}}$$

das heißt: die γ_{ij} treten hier an die Stelle der ν_{ij} in (8.5).

Bei $\alpha_i = 0$ für alle $i \in X$ gehen die γ_{ij} wieder in die ν_{ij} über.

3. Zur zahlenmäßigen Berechnung der Maximum-Likelihood-Schätzung verwendet man statt der in (9.6) angegebenen Formeln für $\gamma(j, k)$, $\gamma^{(i)}(j, k)$, p_k besser Rekursionsformeln in $k \in \mathbb{N}$

4. Es wurde folgendes Beispiel auf einer elektronischen Rechenanlage durchgerechnet:

$X = \{1, 2, 3\}$, Stichprobe x^n vom Umfang $n = 20$:

1112 1121 1221 1232 3332

Es ergaben sich die Maximum-Likelihood-Schätzungen $\hat{\vartheta} = (\hat{\alpha}, \hat{\Lambda})$:

$\hat{\alpha}_1 = 0.826$, $\hat{\alpha}_2 = 0.654$, $\hat{\alpha}_3 = 0.000$

$$(\hat{\Lambda}_{ij}) = \begin{pmatrix} 0.000 & 1.000 & 0.000 \\ 0.778 & 0.000 & 0.222 \\ 0.000 & 0.414 & 0.586 \end{pmatrix}$$

Die in der Einführung erwähnte Charakterisierung der Ereignisse $i \in X$ durch die Koeffizienten α_i findet man in diesem Beispiel bestätigt.

Die Likelihoodfunktion $\frac{1}{n} \log_e p(s, x^n, \vartheta)$ mit Startvektor $s = (1, 0, 0)$ und Parameter $\vartheta = ((\alpha_i), (\Lambda_{ij}))$ wurde für zwei ϑ -Werte berechnet, einmal für die Maximum-Likelihood-Schätzung $\hat{\vartheta}$ (lineare OM-Kette) und einmal für ein ϑ mit allen $\alpha_i = 0$ und $\Lambda_{ij} = \nu_{ij} / \nu_i$ (Markovkette):

Lineare OM-Kette: -0.681 Markovkette: -0.752

Erwartungsgemäß ist der erste Wert größer als der zweite.

Desweiteren weiteren wurde die in Satz (8.8) getroffene Konvergenzaussage für die Maximum-Likelihood-Schätzungen der endlichen linearen OM-Kette rechnerisch nachvollzogen. Die oben errechnete Schätzung $\vartheta = (\hat{\alpha}, \hat{\Lambda})$ wurde als "wahrer" Parameter ϑ_0 gewählt. Mit diesem ϑ_0 wurden mit Hilfe der Monte-Carlo-Methode (vgl. /13/, § 2.3.1.4) nacheinander Stichproben vom Umfang 300, 600, 900 usw, simuliert. Für jede simulierte Stichprobe wurde die Maximum-Likelihood-Schätzung für ϑ berechnet und mit ϑ_0 verglichen. Die Differenz in den einzelnen Kompo-

nenten überstieg erst ab $n = 2100$ nicht den Wert 0.1 und erst ab $n = 3000$ nicht den Wert 0.01 . Natürlich kann in jedem Einzelfall "Scheinkonvergenz" vorliegen. Dieses und weitere durchgerechnete Beispiele warnen davor, bei zu geringen Stichprobenumfängen auf eine gute Übereinstimmung der Maximum-Likelihood-Schätzung mit dem zugrundeliegenden Parameterwert $\hat{\theta}_0$ zu hoffen - namentlich bei einem X mit größerer Anzahl von Elementen.

5. Satz (9.7) gilt, nur leicht modifiziert, auch für die in Bemerkung (4.6) 3. angeführte Erweiterung der endlichen linearen OM-Kette.

ANHANG

Beweis zu Satz (2.7)

Für festes $x \in X$ setzen wir $U\mu := u(\mu, x) \in M$ für alle $\mu \in M$. Ferner bedeute $\sigma(A) := \{\emptyset, A, \mathbb{C}A, X\} \subset \mathfrak{A}$ für alle $A \in \mathfrak{A} (\neq \emptyset, X)$, $I := [0, 1]$

(1) Für jedes $A \in \mathfrak{A} (\neq \emptyset, X)$ existiert eine Funktion $f_A: I \rightarrow I$ mit $U\mu(A) = f_A(\mu(A))$.

Zum Beweis nehme man an, für $\mu, \nu \in M$ gelte $\mu(A) = \nu(A)$. Mit $\mathcal{U} = \sigma(A)$ liefert dann (A_3) , daß auch $U\mu(A) = U\nu(A)$, was aber bedeutet, daß $U\mu(A)$ eine Funktion $f_A(y)$ von $y = \mu(A)$ allein ist. Für jedes $y \in I$ kann $f_A(y)$ festgelegt werden, weil es zu $y \in I$ ein $\mu \in M$ gibt mit $\mu(A) = y$ (Man beachte $A \neq \emptyset, X$).

(2) Für alle $A, B \in \mathfrak{A} (\neq \emptyset, X)$ gilt $f_A(y) - f_A(0) = f_B(y) - f_B(0)$. Wähle zum Beweis A, B mit $A \cap B = \emptyset$ (keine Einschränkung) und $\mu, \nu \in M$ mit $\mu(A+B) = \nu(A+B)$ und $\nu(B) = 0$, also $\mu(A+B) = \nu(A)$. Mit $\mathcal{U}' = \sigma(A+B)$ liefert (A_3) , daß auch $U\mu(A+B) = U\nu(A+B)$, bzw $U\mu(A) + U\mu(B) = U\nu(A) + U\nu(B)$. Setzt man (1) und die Voraussetzungen über μ, ν in diese Gleichung ein, so bekommt man

$$(*) \quad f_A(\mu(A)) + f_B(\mu(B)) = f_A(\mu(A)+\mu(B)) + f_B(0).$$

Wählt man $\mu(A) = 0$ und $\mu(B) = y$, $y \in I$, so ist (2) bewiesen. Über $f_{\emptyset}(0) = 0$, $f_X(1) = 1$ läßt sich mit Hilfe von (2) nun auch $f_{\emptyset}(y)$ und $f_X(y)$ festlegen.

(3) Setze $f(y) := f_A(y) - f_A(0)$. Dann gilt $f(y+z) = f(y) + f(z)$ für $y, z \in I$ und $y+z \in I$.

Denn setzt man in (*) die Def. von $f(y)$ ein, so erhält man $f(\mu(A)) + f_A(0) + f(\mu(B)) + f_B(0) = f(\mu(A)+\mu(B)) + f_A(0) + f_B(0)$. Mit $\mu(A) = y$, $\mu(B) = z$ folgt die Behauptung.

(4) Es gilt $f(y) = \alpha y$ mit $|\alpha| \leq 1$.

Zum Nachweis setze man $\alpha := f(1)$ und beweise $f(y) = \alpha y$ nacheinander für $1/y \in \mathbb{N}$, $y \in \mathbb{Q}$, $y \in \mathbb{R}$ (letzteres wegen der Beschränktheit von f)

(5) Mit $|\alpha| \leq 1$ und $\Lambda \in M$ gilt $U\mu = \alpha\mu + (1-\alpha)\Lambda$.

Die Punkte (1) - (4) haben $U\mu(A) = \alpha\mu(A) + f_A(o)$ für $\mu \in M$, $A \in \mathcal{K}$.^{gebracht}

Es folgt, daß $f_A(o)$ ein endliches Maß auf $\mathcal{K}$ ist mit $f_X(o) = 1-\alpha$.

Für $\alpha = 1$ ist $f_X(o) \equiv 0$ und damit (5) bewiesen. Für $\alpha \neq 1$ setzt man $\Lambda(A) := f_A(o)/(1-\alpha)$ und erhält ebenfalls (5).

(6) Sei $|X| = N$, $\mathcal{K} = \mathcal{J}^O(X)$. Es gibt ein $A \in \mathcal{K}$, $A \neq X$, mit $\Lambda(A) \geq (N-1)/N$. Wähle ein $\mu \in M$ mit $\mu(A) = 0$. Dann folgt aus $\alpha\mu + (1-\alpha)\Lambda \leq 1$ durch Einsetzen dieses A : $\alpha \geq -\frac{1}{N-1}$.

Sei $|X|$ und $|\mathcal{K}| = \infty$. Dann folgt wie eben sofort $\alpha \geq 0$.

Ein entscheidender Punkt in diesem Beweis ist natürlich die Tatsache, daß man die Menge M aller W.maße auf $\mathcal{K}$ als Zustandsraum gewählt hat. Man kommt allerdings mit echten Teilmengen von M aus, z.B. mit der Menge M_ℓ aller W.maße, die bezügl. eines atomlosen, σ -endlichen Maßes ℓ stetig sind, also eine ℓ -Dichte haben.

Wegen der eindeutigen Fortsetzbarkeit eines (σ -endlichen Prä-) Maßes von einer erzeugenden Algebra $\mathcal{A}^*$ auf $\mathcal{K} = \sigma(\mathcal{A}^*)$ genügt es, die Bedingung (A_3) in Satz (2.7) nur für alle Unter-Algebren $\mathcal{A} \subset \mathcal{A}^*$ zu fordern (anstatt für alle Unter- σ -Algebren $\mathcal{A} \subset \mathcal{K}$).

Beweis zu Satz (3.10)

Nach Definition (3.5) existiert eine Nullfolge ε_n mit

$$(*) \left\{ \left| \bar{P}^n(s, A) - P^\infty(A) \right| \leq \varepsilon_n \quad \text{für alle } s \in S, n \in \mathbb{N} \right\}$$

und für alle $A \in \mathfrak{A}^1$, $1 \in \mathbb{N}$. Dabei ist zur Abkürzung gesetzt:

$$\bar{P}^n(s, A) = \frac{1}{n} \sum_{k=1}^n P^k(s, A) .$$

Zuerst zeigen wir, daß es für jedes $A \in \mathfrak{A}^{\mathbb{N}}$ eine Zahl $P^\infty(A)$ gibt, so daß die Ungleichung (*) gilt. (Man beachte, daß $\bar{P}^n(s, A)$ für jedes s und n ein W.maß auf $\mathfrak{A}^{\mathbb{N}}$ darstellt gemäß (1.7)).

Aufgrund des Satzes über "monotone Klassen" (vgl /9/, p.12, Prop.17) genügt es zu zeigen:

Sei $\mathcal{A} \subset \mathfrak{A}^{\mathbb{N}}$ ein Teilsystem mit der Eigenschaft, daß es zu jedem $A \in \mathcal{A}$ eine Zahl $P^\infty(A)$ gibt, so daß (*) gilt. Dann gibt es auch zu jedem $A = \bigcup_{i=1}^{\infty} A_i$, $A_i \in \mathcal{A}$, $A_1 \subset A_2 \subset \dots$, eine Zahl $P^\infty(A)$, so daß (*) gilt.

Dies wird so bewiesen. Für jedes $s \in S$ und $n \in \mathbb{N}$ ist die Zahlenfolge $\bar{P}^n(s, A_i)$ mit wachsendem i monoton nicht fallend und beschränkt, weshalb diese Eigenschaft wegen (*) auch die Zahlenfolge $P^\infty(A_i)$ besitzt. $P^\infty(A)$ bezeichne $\lim_{n \rightarrow \infty} P^\infty(A_i)$. Sei nun n und s fest gewählt. Zu $\varepsilon > 0$ gibt es wegen der Stetigkeit des Maßes $\bar{P}^n(s, \cdot)$ und wegen der Konvergenz $P^\infty(A_i) \rightarrow P^\infty(A)$ ein $n_0 \in \mathbb{N}$ mit

$$\left| \bar{P}^n(s, A_i) - \bar{P}^n(s, A) \right| \leq \varepsilon$$

$$\left| P^\infty(A_i) - P^\infty(A) \right| \leq \varepsilon$$

für alle $i \geq n_0$. Es folgt daraus wegen $A_i \in \mathcal{A}$

$$\left| \bar{P}^n(s, A) - P^\infty(A) \right| \leq \varepsilon_n + 2\varepsilon , \text{ was das erste Beweisziel war.}$$

Die Gleichmäßigkeit der Konvergenz in (*) garantiert dann, daß

$$P^\infty(A) = \lim_n \frac{1}{n} \sum_{k=1}^n P^k(s, A) , A \in \mathfrak{A}^{\mathbb{N}} \text{ ein W.maß auf } \mathfrak{A}^{\mathbb{N}} \text{ ist.}$$

ZUSAMMENFASSUNG

Gegeben sei eine Kette mit vollständigen Bindungen (Kette mvB) mit Ereignisraum $(X, \mathcal{E})$, Zustandsraum $(S, \mathcal{S})$, Übergangswahrscheinlichkeiten $P(s, A)$, $A \in \mathcal{E}$, und Zustandstransformationen $u(s, x)$ (vgl. Def. (1.1)). Die Größen $u(s, x)$ und $P(s, A)$ mögen noch von einem unbekannten Parameter ϑ abhängen. Es wird weiter vorausgesetzt, daß die Wahrscheinlichkeiten $P(s, A)$ bezüglich eines Wahrscheinlichkeitsmaßes auf $\mathcal{E}$ Dichten $p(s, x)$ besitzen, die dann wiederum von ϑ abhängen. Bei vorgegebenem Startzustand s läßt sich dann für eine Stichprobe $x^n = (x_1, \dots, x_n)$ eine Likelihoodfunktion $p(s, x^n, \vartheta)$ angeben. $p(s, x^n, \vartheta)$ nehme im (nicht notwendig eindeutigen) Punkt $\hat{\vartheta}_n$ sein Maximum über ϑ an. Die vorliegende Arbeit behandelt nun die Frage nach der Konsistenz der Maximum-Likelihood-Schätzung $\hat{\vartheta}_n$: Unter welchen Voraussetzungen an die Kette mvB konvergiert $\hat{\vartheta}_n$ bei $n \rightarrow \infty$ P_s -fast sicher gegen den wahren Parameter ϑ_0 ? Dabei ist P_s die gemäß Satz (1.4) zum Startzustand s und zum Parameter ϑ_0 gehörende Wahrscheinlichkeit auf dem Produktraum $(X^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}})$. Das Hauptergebnis ist Satz (7.15). Zu den Voraussetzungen gehören u.a., daß die Transformationen $u(s, x)$ in gewisser Weise kontrahierend sind (vgl. $(B_3), (B_4)$ in §3), und daß die Kette mvB gleichmäßig ergodisch ist (vgl. Def. (3.5)).

Ein konkretes Beispiel einer Kette mvB ist die lineare OM-Kette (2.6), die als Spezialfall die Markovkette enthält. Alle Bedingungen, die in der Arbeit formuliert werden, werden darauf hin geprüft, ob bzw. unter welchen nachprüfbaren Voraussetzungen sie von der linearen OM-Kette erfüllt werden.

Der Inhalt der einzelnen Paragraphen soll nun zusammengefaßt dargestellt werden.

§1 Die grundlegenden Definitionen und Sätze zur Theorie der Ketten mvB, soweit sie für unsere Fragestellung von Bedeutung sind, werden aufgeführt. In Form und Inhalt haben wir uns dabei dem Konzept von Iosifescu und Theodorescu /13/ angeschlossen.

§2 Die in dieser Arbeit als Standardbeispiel dienenden linearen OM-Ketten werden mit Hilfe dreier Axiome aus der allgemeinen Kette mvB hergeleitet. Als Zwischenstufe dient dabei die OM-Kette (2.1). Axiom (A_1) sollte für einen in der Natur vorkommenden stochastischen Prozeß stets gültig sein, (A_2) ist mehr technischer Natur. Axiom (A_3) wurde zuerst von Bush, Mosteller, Thompson für endliches X angegeben. Im Zustandsraum S führen wir eine Metrik ein und zeigen, daß dann die Meßbarkeitsbedingungen der Definition (1.1) erfüllt sind, d.h., daß die lineare OM-Kette tatsächlich eine Kette mvB ist.

§3 Die wichtigste Voraussetzung für den Beweis der Konsistenz der Maximum-Likelihood-Schätzung ist die Ergodizität der Kette mvB. Zur Herleitung der Sätze (3.10), (3.14), (3.17) benötigen wir sie sogar in der Form der gleichmäßigen Ergodizität (3.5). Wir schließen uns hier dem Konzept von Norman /17/ an, dessen Begriff "ergodisch" wir aber "mischend" nennen wollen. An Stelle der Eigenschaft "mischend" (Konvergenz der n-Schritt-Wahrscheinlichkeiten $P^n(s,A)$ gegen eine stationäre Wahrscheinlichkeit $P(A)$) interessieren wir uns mehr für die allgemeinere Eigenschaft "ergodisch" (Konvergenz von $1/n \sum_{k=0}^{n-1} P^k(s,A)$ gegen $P(A)$), weil dadurch auch periodische Prozesse eingeschlossen sind. Die Bedingungen (B_1) bis (B_5), wobei (B_1) bis (B_4) von Norman stammen und von ihm als "distance diminishing model" bezeichnet wurden, und (B_5) eine Erweiterung gegenüber einer ähnlichen Bedingung von Norman ist, garantieren die gleichmäßige Ergodizität der Kette mvB (Satz (3.9)). Die für uns wichtigste Folgerung aus der gleichmäßigen Ergodizität ist Satz (3.17). Er besagt, daß ein terminales Ereignis, dessen stationäre Wahrscheinlichkeit Null ist, auch eine P_g -Nullmenge ist.

Die Bedingungen (B_2) bis (B_4) sind ohne weiteres von den linearen OM-Ketten erfüllbar, (B_1) jedoch nur unter einigen Einschränkungen (3.1).

§4 Einfache Voraussetzungen, unter denen eine endliche lineare OM-Kette die Bedingung (B_5) und damit die gleichmäßige Ergozität erfüllt, werden in Satz (4.5) angegeben.

§5 Hier wird nun angenommen, daß bezüglich eines Wahrscheinlichkeitsmaßes auf $\mathcal{X}$ die Wahrscheinlichkeiten $P(s,A)$ Dichten $p(s,x)$ besitzen. Das Ziel der §§ 5 und 6 ist es, für die Likelihoodfunktion $p(s,x^n)$ einer Stichprobe $x^n = (x_1, \dots, x_n)$ eine Version des Shannon-McMillan Theorems zu beweisen, nämlich die P_s -fast sichere Konvergenz von $1/n \log p(s,x^n)$. An die Dichten $p(s,x^n)$ stellen wir die Bedingungen (C_1) , (C_2) und (C_3) . (C_2) ist dabei analog zu (B_2) und (C_3) stellt eine Bedingung an die großen Werte von $p(s,x)$ dar. Aus dem Beweis des Shannon-McMillan Theorems (vgl. etwa [6], [15]) übernehmen wir die Idee, die stationäre Wahrscheinlichkeit $P|_{\mathcal{X}^{\mathbb{N}}}$ auf $\mathcal{X}^{\mathbb{Z}}$ zu erweitern und neben der Funktion $p(s,x^n)$ eine Funktion $p((\dots, x_{-1}, x_0), x^n)$ zu betrachten, auf welche dann das individuelle Ergodentheorem anwendbar ist. Im §5 werden die technischen Hilfsmittel für diese Ausweitung bereitgestellt.

§6 Die P_s -fast sichere Konvergenz von $1/n \log p(s,x^n)$ wird bewiesen. (Satz (6.5)). Ein wichtiger Teil des Beweises besteht darin, mit Hilfe von Satz (3.17) von der P -fast sicheren Konvergenz auf die P_s -fast sichere Konvergenz zu schließen.

§7 Erst in diesem Paragraphen berücksichtigen wir in der Notation die Abhängigkeit der Kette mvB von einem Parameter ϑ . Die Ergebnisse der §§ 5 und 6 lassen sich mit Hilfe der Bedingungen (D_1) bis (D_3) , die auf die spezielle Art der ϑ -Abhängigkeit eingehen, auf die Funktionen $p(s, x^n, \vartheta)$ und $\{ \sup p(s, x^n, \vartheta') \}$, $\vartheta' \in U(\vartheta) \}$ anstelle von $p(s, x^n)$ übertragen. Dabei sind die Bedingungen $(D_1), (D_2)$ unproblematisch und in Untersuchungen dieser Art üblich, die Bedingung (D_3) dagegen ist komplizierterer Natur. Satz (7.15) stellt dann das Hauptergebnis der statistischen Untersuchung dar. Er behauptet (zusammen mit Bemerkung (7.16)) unter den Bedingungen "gleichmäßige Ergodizität, $(B_1), (C_1), (D_1)$ " die Konsistenz der Maximum-Likelihood-Schätzung, und zwar in der P_s -Wahrscheinlichkeit.

§8 Für eine endliche lineare OM-Kette läßt sich in vollständiger Weise beschreiben, unter welchen Voraussetzungen über ihre Parameter die oben aufgeführten Bedingungen erfüllt sind. Projiziert man die aufgeführten Bedingungen auf einen Markovprozeß, der ja einen Spezialfall der linearen OM-Kette darstellt, so ergeben sich im wesentlichen die gleichen Voraussetzungen für den Konsistenzbeweis wie bei Roussas /23/ oder Gänssler /11/ (vgl. Bemerkung (8.4)). Im Fall der endlichen Markovkette sind dann die benötigten Voraussetzungen auch identisch mit denen bei Billingsley /5/.

§9 In Anlehnung an /3/ wird für die endliche lineare OM-Kette ein Iterationsverfahren zur Berechnung der Maximumstelle $\hat{\vartheta}_n$ von $p(s, x^n, \vartheta)$ entwickelt. An einem konkreten Rechenbeispiel wird das Konvergenzverhalten von $\hat{\vartheta}_n$ demonstriert.

LITERATURVERZEICHNIS

- / 1/ Bauer H. Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie. Berlin 1968.
- / 2/ Baum L.E. und Petrie T. Statistical Inference for Probabilistic Functions of Finite State Markov Chains. Ann. Math. Statistics 37, 1554-1563 (1966).
- / 3/ Baum L.E., Petrie T., Soules G. und Weiss N. A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. Ann. Math. Statistics 41, 164-171 (1970).
- / 4/ Bert M.C. Convergence d'un processus à liaisons complètes, et application à un processus d'apprentissage linéaire. Ann. Inst. H. Poincaré 4, 1-24 (1968).
- / 5/ Billingsley P. Statistical Inference for Markov Processes. The University of Chicago Press 1961.
- / 6/ Billingsley P. Ergodic Theory and Information. New York, London, Sydney 1965.
- / 7/ Bush R.R., Mosteller F. und Thompson G.L. A Formal Structure for Multiple-Choice Situations. In: Decision processes. Eds: Thralls, Coombs, Davis. New York 1954.
- / 8/ Cox D.R. und Miller H.D. The Theory of Stochastic Processes. London 1965.
- / 9/ Dinculeanu N. Vector Measures. Berlin 1966.
- /10/ Feller W. An Introduction to Probability Theory and its Applications. (2nd Edition) Vol. I, New York, London, Sydney 1957.
- /11/ Gänssler P. Note on Minimum Contrast Estimates for Markov Processes. Metrika 19, 115-130 (1972).
- /12/ Hewitt E. und Stromberg K. Real and Abstract Analysis. Berlin, Heidelberg, New York 1969.
- /13/ Iosifescu M. und Theodorescu R. Random Processes and Learning. New York 1969.
- /14/ Jamison B. Asymptotic Behaviour of Successive Iterates of Continuous Functions under a Markov Operator. Journ. Math. Anal. and Appl. 9, 203-214 (1964).

- /15/ Moy, S.C. Generalizations of Shannon-McMillan Theorem. Pacific J.Math. 11, 705-714 und 1459-1465 (1961).
- /16/ Neveu J. Bases mathématiques du calcul des probabilités. Paris 1964.
- /17/ Norman M.F. Some Convergence Theorems for Stochastic Learning Models with Distance Diminishing Operators. Journ. Math. Psychology 5, 61-101 (1968).
- /18/ Norman M.F. Markov Processes and Learning Models. New York und London 1972.
- /19/ Onicescu O. und Mihoc G. Sur les chaînes de variables statistiques. Bull. Sci. Math. 59, 174-192 (1935).
- /20/ Petrie T. Probabilistic Functions of Finite State Markov Chains. Ann. Math. Statistics 40, 97-115 (1969).
- /21/ Pfanzagl J. On the Measurability and Consistency of Minimum Contrast Estimates. Metrika 14, 249-272 (1969).
- /22/ Rao M.M. Inference in Stochastic Processes. II. Z. Wahrscheinlichkeitstheorie 5, 317-335 (1966).
- /23/ Roussas G.G. Extension to Markov Processes of a Result by A. Wald about the Consistency of the Maximum Likelihood Estimate. Z. Wahrscheinlichkeitstheorie 4, 69-73 (1965).
- /24/ Roussas G.G. Asymptotic Normality of the Maximum Likelihood Estimate in Markov Processes. Metrika 14, 62-70 (1969).
- /25/ Theiler G. Uniformly Ergodic Stochastic Processes with Discrete Parameter. Rev. Roum. math. pures et appl. 12, 191-223 (1967).
- /26/ Wald A. Asymptotic Properties of the Maximum Likelihood Estimate of an Unknown Parameter of a Discrete Stochastic Process. Ann. Math. Statistics 19, 40-46 (1948).
- /27/ Wald A. Note on the Consistency of the Maximum Likelihood Estimate. Ann. Math. Statistics 20, 595-601 (1949).
- /28/ Weiß L. Asymptotic Properties of Maximum Likelihood Estimators in some Nonstandard Cases. Journ. American Stat. Ass. 66, 345-350 (1971).
- /29/ Yosida K. Functional Analysis. (2nd edition). Berlin, Heidelberg, New York 1968.

LEBENS LAUF

Am 13. Juni 1943 wurde ich als zweites von drei Kindern des Spediteurs Friedrich Pruscha und seiner Ehefrau Aloisia geb. Höhnel in Teplitz-Schönau im Sudetenland geboren. Nach der Vertreibung im Jahre 1945 siedelte unsere Familie nach Wetter an der Ruhr über. Dort besuchte ich ab Ostern 1949 die Volksschule und ab Ostern 1954 das Gymnasium. Im Februar 1963 legte ich meine Reifeprüfung am Leibniz-Gymnasium in Düsseldorf ab. Im Sommersemester 1963 begann ich mit dem Studium der Theologie und nach zwei Semestern mit dem Studium der Mathematik an der Universität Bonn. Ab Wintersemester 1965/66 studierte ich an der Universität Freiburg im Breisgau und ab Wintersemester 1966/67 an der Universität München. Nachdem ich im Februar 1969 die Diplomhauptprüfung in Mathematik ablegte, war ich ein halbes Jahr lang wissenschaftliche Hilfskraft bei Prof. Richter. Seit August 1969 bin ich als wissenschaftlicher Mitarbeiter am Max-Planck-Institut für Psychiatrie in München beschäftigt, und zwar in der von Prof. Ploog geleiteten Abteilung für Verhaltensforschung. Ich arbeite dort an den statistischen Auswertungen der von Dr. Maurus durchgeführten Hirnreizexperimente, in denen das Sozialverhalten von Primaten untersucht wird. Daneben wurde mir Gelegenheit geboten, an meiner Dissertation zu arbeiten.

Helmut Pruscha

